

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/42460 A2

(51) International Patent Classification⁷: **C12N 15/12**,
C07K 14/47, C12Q 1/68, G01N 33/68

Lugano, San Jose, CA 95120 (US). **SUN, Yongming**
[US/US]; 869 S. Winchester Boulevard, Apartment 260,
San Jose, CA 95128 (US). **LIU, Chenghua** [CN/US];
1125 Ranchero Way #14, San Jose, CA 95118 (US).

(21) International Application Number: PCT/US01/43611

(22) International Filing Date:
21 November 2001 (21.11.2001)

(74) Agents: **LICATA, Jane, Massey et al.**; Licata & Tyrell
P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/252,505 22 November 2000 (22.11.2000) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/252,505 (CIP)
Filed on 22 November 2000 (22.11.2000)

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): **DI-
ADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard,
South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **MACINA, Roberto**,
A. [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136
(US). **RECIPON, Herve** [FR/FR]; 85 Fortuna Avenue,
San Francisco, CA 94115 (US). **PLUTA, Jason** [US/US];
Apartment 15, 1240 Dale Avenue, Mountain View, CA
94040 (US). **GHOSH, Malavika, G.** [IN/US]; 1733 Via

Published:

— without international search report and to be republished
upon receipt of that report

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

WO 02/42460 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO COLON SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic colon cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon tissue, identifying colon tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered colon tissue for treatment and research.

COMPOSITIONS AND METHODS RELATING TO COLON SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application
5 Serial No. 60/252,505 filed November 22, 2000, which is herein incorporated by
reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and
10 polypeptides present in normal and neoplastic colon cells, including fragments, variants
and derivatives of the nucleic acids and polypeptides. The present invention also relates
to antibodies to the polypeptides of the invention, as well as agonists and antagonists of
the polypeptides of the invention. The invention also relates to compositions comprising
the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists
15 of the invention and methods for the use of these compositions. These uses include
identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-
cancerous disease states in colon tissue, identifying colon tissue and monitoring and
identifying and/or designing agonists and antagonists of polypeptides of the invention.
The uses also include gene therapy, production of transgenic animals and cells, and
20 production of engineered colon tissue for treatment and research.

BACKGROUND OF THE INVENTION

Colorectal cancer is the second most common cause of cancer death in the United
States and the third most prevalent cancer in both men and women. M. L. Davila & A.
D. Davila, *Screening for Colon and Rectal Cancer*, in Colon and Rectal Cancer 47 (Peter
25 S. Edelstein ed., 2000). Approximately 100,000 patients every year suffer from colon
cancer and approximately half that number die of the disease. Hannah-Ngoc Ha & Bard
C. Cosman, *Treatment of Colon Cancer*, in Colon and Rectal Cancer 157 (Peter S.
Edelstein ed., 2000). Nearly all cases of colorectal cancer arise from adenomatous
polyps, some of which mature into large polyps, undergo abnormal growth and
30 development, and ultimately progress into cancer. Davila & Davila, *supra* at 55-56.
This progression would appear to take at least 10 years in most patients, rendering it a

readily treatable form of cancer if diagnosed early, when the cancer is localized. *Id.* at 56; Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 125 (1998).

Although our understanding of the etiology of colon cancer is undergoing continual refinement, extensive research in this area points to a combination of factors, including age, hereditary and nonhereditary conditions, and environmental/dietary factors. Age is a key risk factor in the development of colorectal cancer, Davila & Davila, *supra* at 48, with men and women over 40 years of age become increasingly susceptible to that cancer, Burdette, *supra* at 126. Incidence rates increase considerably in each subsequent decade of life. Davila et al., *supra* at 48. A number of hereditary and nonhereditary conditions have also been linked to a heightened risk of developing colorectal cancer, including familial adenomatous polyposis (FAP), hereditary nonpolyposis colorectal cancer (Lynch syndrome or HNPCC), a personal and/or family history of colorectal cancer or adenomatous polyps, inflammatory bowel disease, diabetes mellitus, and obesity. *Id.* at 47; Henry T. Lynch & Jane F. Lynch, *Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndromes)*, in Colon and Rectal Cancer 67-68 (Peter S. Edelstein ed., 2000).

In the case of FAP, the tumor suppressor gene APC (adenomatous polyposis coli), located at 5q21, has been either mutationally inactivated or deleted. Alberts et al., Molecular Biology of the Cell 1288 (3d ed. 1994). The APC protein plays a role in a number of functions, including cell adhesion, apoptosis, and repression of the *c-myc* oncogene. N. R. Hall & R. D. Madoff, *Genetics and the Polyp-Cancer Sequence*, Colon and Rectal Cancer 8 (Peter S. Edelstein, ed., 2000). Of those patients with colorectal cancer who have normal APC genes, over 65% have such mutations in the cancer cells but not in other tissues. Alberts et al., *supra* at 1288. In the case of HNPCC, patients manifest abnormalities in the tumor suppressor gene HNPCC, but only about 15% of tumors contain the mutated gene. *Id.* A host of other genes have also been implicated in colorectal cancer, including the *K-ras*, *N-ras*, *H-ras* and *c-myc* oncogenes, and the tumor suppressor genes *DCC* (deleted in colon carcinoma) and *p53*. Hall & Madoff, *supra* at 8-9; Alberts et al., *supra* at 1288.

Environmental/dietary factors associated with an increased risk of colorectal cancer include a high fat diet, intake of high dietary red meat, and sedentary lifestyle. Davila & Davila, *supra* at 47; Reddy, B. S., *Prev. Med.* 16(4): 460-7 (1987).

Conversely, environmental/dietary factors associated with a reduced risk of colorectal cancer include a diet high in fiber, folic acid, calcium, and hormone-replacement therapy in post-menopausal women. Davila & Davila, *supra* at 50-55. The effect of antioxidants in reducing the risk of colon cancer is unclear. *Id.* at 53.

5 Because colon cancer is highly treatable when detected at an early, localized stage, screening should be a part of routine care for all adults starting at age 50, especially those with first-degree relatives with colorectal cancer. One major advantage of colorectal cancer screening over its counterparts in other types of cancer is its ability to not only detect precancerous lesions, but to remove them as well. Davila & Davila,
10 *supra* at 56. The key colorectal cancer screening tests in use today are fecal occult blood test, sigmoidoscopy, colonoscopy, double-contrast barium enema, and the carcinoembryonic antigen (CEA) test. *Id.*; Burdette, *supra* at 125.

 The fecal occult blood test (FOBT) screens for colorectal cancer by detecting the amount of blood in the stool, the premise being that neoplastic tissue, particularly
15 malignant tissue, bleeds more than typical mucosa, with the amount of bleeding increasing with polyp size and cancer stage. Davila & Davila, *supra* at 56-57. While effective at detecting early stage tumors, FOBT is unable to detect adenomatous polyps (premalignant lesions), and, depending on the contents of the fecal sample, is subject to rendering false positives. *Id.* at 56-59. Sigmoidoscopy and colonoscopy, by contrast,
20 allow direct visualization of the bowel, and enable one to detect, biopsy, and remove adenomatous polyps. *Id.* at 59-60, 61. Despite the advantages of these procedures, there are accompanying downsides: sigmoidoscopy, by definition, is limited to the sigmoid colon and below, colonoscopy is a relatively expensive procedure, and both share the risk of possible bowel perforation and hemorrhaging. *Id.* at 59-60. Double-contrast barium
25 enema (DCBE) enables detection of lesions better than FOBT, and almost as well as colonoscopy, but it may be limited in evaluating the winding rectosigmoid region. *Id.* at 60. The CEA blood test, which involves screening the blood for carcinoembryonic antigen, shares the downside of FOBT, in that it is of limited utility in detecting colorectal cancer at an early stage. Burdette, *supra* at 125.

30 Once colon cancer has been diagnosed, treatment decisions are typically made in reference to the stage of cancer progression. A number of techniques are employed to stage the cancer (some of which are also used to screen for colon cancer), including

pathologic examination of resected colon, sigmoidoscopy, colonoscopy, and various imaging techniques. AJCC Cancer Staging Handbook 84 (Irvin D. Fleming et al. eds., 5th ed. 1998); Montgomery, R. C. and Ridge, J.A., *Semin. Surg. Oncol.* 15(3): 143-150 (1998). Moreover, chest films, liver functionality tests, and liver scans are employed to
5 determine the extent of metastasis. Fleming et al. eds., *supra* at 84. While computerized tomography and magnetic resonance imaging are useful in staging colorectal cancer in its later stages, both have unacceptably low staging accuracy for identifying early stages of the disease, due to the difficulty that both methods have in (1) revealing the depth of bowel wall tumor infiltration and (2) diagnosing malignant adenopathy. Thoeni, R. F.,
10 *Radiol. Clin. N. Am.* 35(2): 457-85 (1997). Rather, techniques such as transrectal ultrasound (TRUS) are preferred in this context, although this technique is inaccurate with respect to detecting small lymph nodes that may contain metastases. David Blumberg & Frank G. Opelka, *Neoadjuvant and Adjuvant Therapy for Adenocarcinoma of the Rectum, in Colon and Rectal Cancer* 316 (Peter S. Edelstein ed., 2000).

15 Several classification systems have been devised to stage the extent of colorectal cancer, including the Dukes' system and the more detailed International Union against Cancer-American Joint Committee on Cancer TNM staging system, which is considered by many in the field to be a more useful staging system. Burdette, *supra* at 126-27. The TNM system, which is used for either clinical or pathological staging, is divided into four
20 stages, each of which evaluates the extent of cancer growth with respect to primary tumor (T), regional lymph nodes (N), and distant metastasis (M). Fleming et al. eds., *supra* at 84-85. The system focuses on the extent of tumor invasion into the intestinal wall, invasion of adjacent structures, the number of regional lymph nodes that have been affected, and whether distant metastasis has occurred. *Id.* at 81.

25 Stage 0 is characterized by *in situ* carcinoma (Tis), in which the cancer cells are located inside the glandular basement membrane (intraepithelial) or lamina propria (intramucosal). *Id.* at 84-85; Burdette, *supra* at 127. In this stage, the cancer has not spread to the regional lymph nodes (N0), and there is no distant metastasis (M0). Fleming et al. eds., *supra* at 85; Burdette, *supra* at 127. In stage I, there is still no spread
30 of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the submucosa (T1) or has progressed further to invade the muscularis propria (T2). Fleming et al. eds., *supra* at 84-85; Burdette, *supra* at 127. Stage II also involves

no spread of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the subserosa, or the nonperitonealized pericolic or perirectal tissues (T3), or has progressed to invade other organs or structures, and/or has perforated the visceral peritoneum (T4). *Id.* Stage 3 is characterized by any of the T substages, no
5 distant metastasis, and either metastasis in 1 to 3 regional lymph nodes (N1) or metastasis in four or more regional lymph nodes (N2). Fleming et al. eds., *supra* at 85; Burdette, *supra* at 127. Lastly, stage 4 involves any of the T or N substages, as well as distant metastasis. *Id.*

Currently, pathological staging of colon cancer is preferable over clinical staging
10 as pathological staging provides a more accurate prognosis. Pathological staging typically involves examination of the resected colon section, along with surgical examination of the abdominal cavity. Fleming et al. eds., *supra* at 84. Clinical staging would be a preferred method of staging were it at least as accurate as pathological staging, as it does not depend on the invasive procedures of its counterpart.

15 Turning to the treatment of colorectal cancer, surgical resection results in a cure for roughly 50% of patients. Burdette, *supra* at 125. Irradiation is used both preoperatively and postoperatively in treating colorectal cancer. *Id.* at 125, 132-33. Chemotherapeutic agents, particularly 5-fluorouracil, are also powerful weapons in treating colorectal cancer. *Id.* at 125, 133. Other agents include irinotecan and
20 floxuridine, cisplatin, levamisole, methotrexate, interferon-alpha, and leucovorin. *Id.* at 133. Nonetheless, thirty to forty percent of patients will develop a recurrence of colon cancer following surgical resection. Wayne De Vos, *Follow-up After Treatment of Colon Cancer, Colon and Rectal Cancer* 225 (Peter S. Edelstein ed., 2000), which in many patients is the ultimate cause of death. Accordingly, colon cancer patients must be
25 closely monitored to determine response to therapy and to detect persistent or recurrent disease and metastasis.

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of colorectal cancer are of critical importance to the outcome of the patient. Moreover, current procedures,
30 while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that

would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop colorectal cancer, for diagnosing
5 colorectal cancer, for monitoring the progression of the disease, for staging the colorectal cancer, for determining whether the colorectal cancer has metastasized, and for imaging the colorectal cancer. There is also a need for better treatment of colorectal cancer.

SUMMARY OF THE INVENTION

10 The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon; identify and monitor colon tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also
15 provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered colon tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to colon cells and/or colon tissue. These colon specific nucleic acids (CSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of
20 one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the CSNA is genomic DNA, then the CSNA is a colon specific gene (CSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 75 through 124. In another
25 highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 74. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding a CSP, or that selectively hybridize or exhibit substantial sequence similarity to a CSNA, as well as allelic variants of a nucleic
30 acid molecule encoding a CSP, and allelic variants of a CSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes a CSP or that comprises a part of a nucleic acid sequence of a CSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a CSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the
5 transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a CSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of a CSP. In another preferred embodiment, the
10 nucleic acid molecule comprises all or a part of a CSNA.

Another object of the invention is to provide methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic
15 acid molecule of the invention. In a preferred embodiment, the polypeptide is a CSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (muten), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a CSP.

Another object of the invention is to provide an antibody that specifically binds to
20 a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical
25 nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the
30 invention for identifying and/or monitoring colon tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered colon tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon. The invention provides methods of using the polypeptides of the invention to identify and/or monitor colon tissue, and to produce engineered colon tissue.

5 The agonists and antagonists of the instant invention may be used to treat colon cancer and non-cancerous disease states in colon and to produce engineered colon tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for
10 comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by
15 those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used
20 in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press (1989) and
25 *Sambrook et al., Molecular Cloning: A Laboratory Manual*, 3d ed., Cold Spring Harbor Press (2001); Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel et al., *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology - 4th Ed.*, Wiley & Sons (1999); Harlow and Lane, *Antibodies: A Laboratory*
30 *Manual*, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

- 10 A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and
- 15 "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.
- 20 The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates,
- 25 phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially
- 30 duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are

known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround
5 the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a
10 nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

15 A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (e.g., an
20 RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, e.g., ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a
25 portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized
30 polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site,

recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome; and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by
5 expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as
10 those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other
15 polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages
20 such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See *e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical
25 Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In
30 addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each

nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

5 The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

10 The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number
15 of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity
20 of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent
25 sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

30 A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its

complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology"

5 interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its
10 complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

15 Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at
20 least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species,
25 length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of
30 hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is

performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence
5 hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where l is the length of the hybrid in base pairs.

10 The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50$$

15 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic
20 acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization
25 conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours).
30 Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten

hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify
5 nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents
10 to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook
15 (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in
20 the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon
25 degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (e.g., for oligonucleotide probes) may be calculated by the formula:
$$T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/\text{N}),$$

wherein N is change length and the $[\text{Na}^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p.
30 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using

mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

5 The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 µg of plasmid or DNA fragment
10 is digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are
15 specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the
20 art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

25 Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon
30 portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-

derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one
5 exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic
10 acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing
15 Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that
20 nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the
25 nucleic acid molecule comprises the wild type nucleic acid sequence encoding a CSP or is a CSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high
30 rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g., Reidhaar-Olson et al., Science 241: 53-57 (1988).*

5 The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

10 The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g., Stemmer, Proc. Natl. Acad. Sci. U.S.A. 91: 10747-10751 (1994).* DNA shuffling can be carried out between several related genes ("Family shuffling").

15 The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate
20 random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

25 The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g., Arkin et al., Proc. Natl. Acad. Sci. U.S.A. 89: 7811-7815*
30 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein

small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. See, e.g., Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in
5 its entirety.

“Operatively linked” expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

10 The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination,
15 promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such
20 control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term “control sequences” is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

25 The term “vector,” as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a “plasmid”, which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of
30 vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are

introduced (e.g., bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors
5 are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that
10 serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding
15 generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in
20 its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

25 As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-
30 occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide

each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a CSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide
5 that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be
10 "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single
15 species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,
20 followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared
25 to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40
30 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a

different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs.

- 5 The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide
- 10 analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

- The term "non-peptide analog" refers to a compound with properties that are
- 15 analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm
- 20 polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH₂NH--, --CH₂S--, --CH₂-CH₂--,
- CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂--, and --CH₂SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a
- 25 consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may
- 30 add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (e.g., a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure,

Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), *Immunology - A Synthesis* 2nd Ed.,
5 Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α -, β -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate,
10 -N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and
15 convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to
20 another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred
25 embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

30 When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence

similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g., Pearson, Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al., Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g., GCG Version 6.1. Other programs include FASTA, discussed supra.*

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. See, e.g., Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein

5 incorporated by reference. Preferred parameters for blastp are:

- | | | |
|----|-----------------------|---------------|
| | Expectation value: | 10 (default) |
| | Filter: | seg (default) |
| | Cost to open a gap: | 11 (default) |
| | Cost to extend a gap: | 1 (default) |
| 10 | Max. alignments: | 100 (default) |
| | Word size: | 11 (default) |
| | No. of descriptions: | 100 (default) |
| | Penalty Matrix: | BLOSUM62 |

The length of polypeptide sequences compared for homology will generally be at
15 least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms
20 other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (e.g., FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be
25 determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding
portion thereof that competes with the intact antibody for specific binding to a molecular
30 species, e.g., a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv,

dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')₂ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. See, e.g., Ward *et al.*, *Nature* 341: 544-546 (1989).

10 By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

15 A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. See, e.g., Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but
20 using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. See e.g., Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it
25 an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one
30 or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For

instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than $1\ \mu\text{M}$, preferably less than $100\ \text{nM}$ and most preferably less than $10\ \text{nM}$.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "colon specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the colon as compared to other tissues in the body. In a preferred embodiment, a "colon specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "colon specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more

- preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

- One aspect of the invention provides isolated nucleic acid molecules that are specific to the colon or to colon cells or tissue or that are derived from such nucleic acid molecules. These isolated colon specific nucleic acids (CSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon, a colon-specific polypeptide (CSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 75 through 124. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 74.

- A CSNA may be derived from a human or from another animal. In a preferred embodiment, the CSNA is derived from a human or other mammal. In a more preferred embodiment, the CSNA is derived from a human or other primate. In an even more preferred embodiment, the CSNA is derived from a human.

- By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a CSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a CSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a CSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 75 through 124. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 74.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a CSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a CSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a CSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 75 through 124. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 74. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a CSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human CSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 75 through 124. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a CSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 75 through 124, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a CSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a CSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a CSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 74. In a preferred
5 embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a CSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 74, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a CSNA, more preferably at least 95%,
10 more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a CSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one
15 that exhibits sequence identity over its entire length to a CSNA or to a nucleic acid molecule encoding a CSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the CSNA or the nucleic acid molecule encoding a CSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at
20 least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 75 through 124 or demonstrates
25 significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 74. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the CSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated
30 species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived

from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In
5 another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a CSNA. Further, the substantially similar nucleic acid molecule may or may not be a CSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a CSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a
10 CSNA or a nucleic acid encoding a CSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population.
15 Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic
20 variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a CSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a CSP comprising an amino acid sequence of SEQ ID NO: 75 through 124. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a CSNA. In a more preferred embodiment, the gene
25 is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 74. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic
30 acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a CSP. However, in a preferred embodiment, the part encodes a CSP. In one aspect, the invention comprises a part of a

CSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a CSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a CSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a CSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutin or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (e.g., reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or

fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as ^{32}P -dATP, ^{32}P -dCTP, ^{32}P -dGTP, ^{32}P -dTTP, ^{32}P -3'dATP, ^{32}P -ATP, ^{32}P -CTP, ^{32}P -GTP, ^{32}P -UTP, ^{35}S -dATP, α - ^{35}S -GTP, α - ^{33}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for

RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable
5 labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label
10 (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); see Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker
15 (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the
20 nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. See, e.g., Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci.*
25 *USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by
30 reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside

bonds. See Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. See Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and

methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4–16°C (11°C on average).

Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. See, e.g.,

- 5 Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

- Nucleic acid molecules may be modified compared to their native structure
- 10 throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24:
- 15 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

- Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed,
- 20 partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by
- 25 reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a CSNA, such as deletions, insertions, translocations, and duplications of the CSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g., Andreeff et al. (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications*, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify CSNA in, and isolate CSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g., Schwarczacher et al., In Situ Hybridization*, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing

nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to CSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the
5 nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in
10 their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a CSP. In a more
15 preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 75 through 124. In another preferred embodiment, the probe or primer is derived from a CSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 74.

20 In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer
25 in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide
30 probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter*

- alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety:

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as

hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes
5 (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

10 Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui
15 (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

20 Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of
25 this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for
30 example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

-46-

- In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the
- 5 nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λ GT10 and λ GT11, and other phages, *e.g.*,
- 10 M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.
- 15 In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components,
- 20 *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEplac series
- 25 plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most
- 30 common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

-47-

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors
5 are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the
10 polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway.
15 Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, e.g., in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A).
20 Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in
25 mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (e.g., vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (e.g., bovine papillomavirus),
30 and retroviral vectors (e.g., murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For
5 example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally
10 contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of
15 the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding
20 proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*,
25 will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further
30 include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast α -mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the

5 transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the

10 Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the CSNA of interest. Often, expression is enhanced by

15 incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker

20 gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted

25 in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product

30 information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the *trc* promoter, which is regulated by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the *PLtetO-1* promoter. The *PLtetO-1* promoter takes advantage of the high expression levels from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity

resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5
5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2
10 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or
15 identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous
20 phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996).
Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA,
25 USA), use the -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

30 A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like

chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. See Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. See Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (see, e.g. Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (see, e.g., Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (see, e.g., Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (see, e.g., Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. See also Conn (ed.), Green Fluorescent Protein (Methods in

Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

5 Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable
10 expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin
15 C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney
20 murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency;
25 varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art
30 may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must

be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present
5 invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,
10 and acylation, and it is an aspect of the present invention to provide CSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation,
15 acetylation, prenylation, methylation, acetylation, arginylation, ubiquitination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational
20 modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-
25 anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in
30 web sites such as the Delta Mass database <http://www.abrf.org/ABRF/ResearchCommittees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological

sources" Cooper et al. *Nucleic Acids Res.* 29; 332-335 (2001) and <http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27: 370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) and <http://www.cbs.dtu.dk/databases/PhosphoBase/> (accessed October 19, 2001); or <http://pir.georgetown.edu/pirwww/search/textresid.html> (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to
5 the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is
10 cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell
15 compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method
20 known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel
25 electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the
30 art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically.

- 5 Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule
- 10 encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired
- 15 post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid
- 20 sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

- In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its
- 25 controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture
- 30 requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly
5 expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors
10 typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation,
15 liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (See, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector
20 comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to
25 express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture.
30 Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces*

- pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical
- 5 mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and
- 10 readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from colon are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human colon cells.
- 15 Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. See, e.g., Ausubel (1992), *supra*, Ausubel
- 20 (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

- 25 Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

- Plasmid vectors will typically be introduced into chemically competent or
- 30 electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, e.g., with CaCl_2 , or a solution of Mg^{2+} , Mn^{2+} , Ca^{2+} , Rb^+ or K^+ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80

- (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (e.g., Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols (BioRad, Richmond, CA, USA) ([http://www.biorad.com/LifeScience/pdf/](http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf)).

- Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

- For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

- For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991).
- The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.,* Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.,* Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by

means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

5 Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a colon specific polypeptide (CSP). In an even more preferred embodiment, the polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ
10 ID NO: 75 through 124. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide,
15 wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a CSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 75 through 124. A polypeptide that comprises only a fragment of an entire CSP may or may not be a polypeptide that is also a CSP. For instance, a full-length polypeptide may be
20 colon-specific, while a fragment thereof may be found in other tissues as well as in colon. A polypeptide that is not a CSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-CSP antibodies. However, in a preferred embodiment, the part or fragment is a CSP. Methods of determining whether a polypeptide is a CSP are described *infra*.

25 Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. See, e.g., Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by
30 native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g.,* Lerner, *Nature* 299: 592-596 (1982); Shinnick *et al., Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al., Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

10 Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

20 The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

25 One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.,* a CSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.,* Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably a CSP, may be produced by

chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a CSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be colon-specific. In a preferred embodiment, the mutein is colon-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 75 through 124. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 75 through 124. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 75 through 124.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino

acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is colon-specific, as described below. Multiple random mutations can be introduced into the

5 gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*,

10 Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the

15 polypeptide is homologous to a CSP. In an even more preferred embodiment, the polypeptide is homologous to a CSP selected from the group having an amino acid sequence of SEQ ID NO: 75 through 124. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a CSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity

20 to an comprising an amino acid sequence of SEQ ID NO: 75 through 124. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 75 through 124. In a yet more

25 preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 75 through 124. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%,

30 99.7%, 99.8% or 99.9% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 75 through 124. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a CSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a CSNA under low stringency, moderate stringency or high stringency conditions, as
5 defined herein. In a more preferred embodiment, the CSNA is selected from the group consisting of SEQ ID NO: 1 through 74. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a CSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the CSP
10 is selected from the group consisting of SEQ ID NO: 75 through 124.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of
15 SEQ ID NO: 75 through 124. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the CSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster,
20 cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous
25 polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a CSP. Further, the
30 homologous protein may or may not encode polypeptide that is a CSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a CSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but
5 also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in
10 eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a CSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino
15 acid sequence selected from the group consisting of SEQ ID NO: 75 through 124. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 74.

In another embodiment, the invention provides polypeptides which comprise
20 derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a CSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 75 through 124, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been
25 acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

30 Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of

glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.),

5 Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of

10 ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide,

15 including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic

20 processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on

25 the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546,

30 Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa

Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

10 The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, 15 HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, 20 SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

25 The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

30 The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-CSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

- 10 In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a CSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 75 through 124. In a preferred embodiment, the analog is one that comprises one or
- 15 more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a CSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂--,
- CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂-- and --CH₂SO--. In another
- 20 embodiment, the non-peptide analog comprises substitution of one or more amino acids of a CSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific
- 25 three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Krole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.
- 30 Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are

- described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entirety.

- Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinyl-(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and tBOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-Fmoc-L-glutamic acid or the corresponding tBOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated Fmoc synthesis of peptides using (Fmoc)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

- Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

- A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, e.g., Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-

- trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-
- 5 2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- β -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-
- 10 hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-
- 15 methylbenzoic acid, Fmoc-3-amino-2-naphthoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-
- 20 1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

- Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed
- 25 mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*,
- 30 *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a CSP. In a more preferred embodiment, the polypeptide
5 that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 75 through 124, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 74, or comprises all or part of a nucleic acid sequence that
10 selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 74.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of
15 the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present
20 invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the
25 present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically
30 synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so

included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of
5 recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG
10 can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins
15 that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8):
20 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-
25 dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbrizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-
30 12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95,

-75-

14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A
5 comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

10 Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be
15 fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, myc, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin trpE, protein A, β -lactamase, -amylase,
20 maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), DNA binding domains, GAL4 DNA binding domains, and the like.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize CSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly CSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of CSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of CSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

-77-

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens
5 to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present
10 invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically
15 derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate,
20 such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when
25 the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with
30 sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so

attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

5 In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies
10 are specific for a polypeptide that is a CSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 75 through 124, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes,
15 discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a CSP may be glycosylated in cancerous cells,
20 but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of a CSP may be indicative of cancer. Differential degradation of the C or N-terminus of a CSP may also be a marker or target for anticancer therapy. For example, a CSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

25 As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-CSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-
30 fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine

the presence of the protein of the present invention in samples derived from human colon.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but

also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are
5 hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids
10 of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments
15 of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725
20 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant,
25 and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in

immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues

5 (Viking et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998)).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody

10 Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal

15 antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two

20 necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

25 Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies

30 of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein

(pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. See, e.g., Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997);
5 Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994).
Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. See, e.g., Barbas
10 (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length
15 antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. See, e.g., Takahashi *et al.*, *Biosci.*
20 *Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3): 157-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997); Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein
25 by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. See, e.g., Li *et al.*, *Protein Expr. Purif.*
21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997);
30 and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,
5 *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entirety.

Antibodies, including antibody fragments and derivatives, of the present
10 invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock *et al.*, *J. Immunol. Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entirety.

Mammalian cells useful for recombinant expression of antibodies, antibody
15 fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

20 Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entirety.

25 The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and
30 protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that
5 bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or
10 one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half
15 life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21):
20 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region,
25 Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody
30 complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium

(INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

- Other substrates can be used to produce products for local deposition that are
- 5 luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic
- 10 compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g., Thorpe et al., Methods Enzymol.* 133: 331-53 (1986); Kricka *et al., J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al., J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such
- 15 enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

- 20 There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein

25 (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY

30 R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY

630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, 5 Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, e.g., for Western blotting 10 applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H , and ^{125}I .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , 15 ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

20 As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the 25 antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

30 The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention,

to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred
5 embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a CSP. In a preferred embodiment, the CSP comprises an amino acid sequence selected from SEQ ID NO: 75 through 124, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a CSNA of the invention, preferably a CSNA
10 comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 74, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human CSG.
15 The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g., Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual*, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al., Mouse Genetics and Transgenics: A Practical Approach*, Oxford University Press (2000); and Pinkert, *Transgenic Animal Technology: A Laboratory Handbook*, Academic Press (1999).
20

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g., Paterson et al., Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al., Biotechnology* 11: 1263-1270 (1993); Wright *et al., Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989) retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g., Van der Putten et al., Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g., Thompson et al., Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g., Lo, 1983, Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g., Ulmer et al., Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem
25
30

-90-

cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g., Lavitrano et al., Cell 57: 717-723 (1989)*).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,*
5 *Campbell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810-813 (1997)*). The present invention provides for transgenic animals that carry the transgene (*i.e., a nucleic acid molecule of the invention*) in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i. e., mosaic animals or chimeric animals.*

The transgene may be integrated as a single transgene or as multiple copies, such
10 as in concatamers, *e. g., head-to-head tandems or head-to-tail tandems.* The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g., the teaching of Lasko et al. et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992).* The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill
15 in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression
20 of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the
25 transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in
30 order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to

both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

- 5 Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.
- 10 Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous
- 15 gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et*
- 20 *al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the

directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

-93-

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 74 and SEQ ID NO: 75 through 124 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data

set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences
5 comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or
10 data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and
15 similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a
20 nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence
25 comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the
30 steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one

overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Colon Cancer

5 The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a CSNA or a CSP in a human patient that has or may have colon cancer, or who is at risk of developing colon cancer, with the expression of a CSNA or a CSP in a normal human control. For purposes of the present invention,
10 “expression of a CSNA” or “CSNA expression” means the quantity of CSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term “expression of a CSP” or “CSP expression” means the amount of CSP that can be measured by any method known in the art or the level of translation of a CSG
15 CSNA that can be measured by any method known in the art.

 The present invention provides methods for diagnosing colon cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of CSNA or CSP in cells, tissues, organs or bodily fluids compared with levels of CSNA or CSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human
20 control, wherein an increase, or decrease in certain cases, in levels of a CSNA or CSP in the patient versus the normal human control is associated with the presence of colon cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in the structure of the mRNA of a CSG compared to the mRNA from a normal
25 control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in a CSP compared to a CSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the CSP or
30 subcellular CSP localization.

 In a preferred embodiment, the expression of a CSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from

SEQ ID NO: 75 through 124, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the CSNA expression that is measured is the level of expression of a CSNA mRNA selected from SEQ ID NO: 1 through 74, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these

5 nucleic acids. CSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.,* Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. CSNA transcription may be

10 measured by any method known in the art including using a reporter gene hooked up to the promoter of a CSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.,* aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, CSNA expression may be compared to a known control, such as normal colon nucleic

15 acid, to detect a change in expression.

In another preferred embodiment, the expression of a CSP is measured by determining the level of a CSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 75 through 124, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs

20 and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of CSNA or CSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of colon cancer. The expression level of a CSP may be determined by any method known in the art, such as those

25 described *supra*. In a preferred embodiment, the CSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.,* Harlow (1999),

30 *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the CSP structure may be determined by any method known in the art, including, *e.g.,* using antibodies that specifically recognize phosphoserine, phosphothreonine or

-97-

phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a CSP is prepared if one is not already available. In a preferred
5 embodiment, the antibody is a monoclonal antibody. The anti-CSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the CSP will bind to the anti-CSP antibody. The sample is removed, the solid support is washed to remove unbound
10 material, and an anti-CSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the CSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the
15 amount of a CSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure CSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-CSP antibody is attached to a solid
20 support and an allocated amount of a labeled CSP and a sample of interest are incubated with the solid support. The amount of labeled CSP detected which is attached to the solid support can be correlated to the quantity of a CSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential
25 separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly
30 square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

-98-

Expression levels of a CSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example,

5 reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

10 Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more CSNAs of interest. In this approach, all or a portion of one or more CSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the
15 RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid
20 molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any
25 other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of CSNA or CSP includes, without limitation, colon tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, colon cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment,
30 especially when metastasis of a primary colon cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands

limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in
5 CSNAs or CSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a CSNA or CSP. In many cases, the use of another cancer marker
10 will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other CSNA or CSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular CSNA or CSP is measured. In a more
15 preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression
20 levels and/or structural alterations of one or more CSNAs and/or CSPs in a sample from a patient suspected of having colon cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP and then ascertaining whether the patient has colon cancer from the expression level of the CSNA or CSP. In general, if high expression
25 relative to a control of a CSNA or CSP is indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is
30 indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the

-100-

same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether colon cancer has metastasized in a patient. One may identify whether the colon cancer has

5 metastasized by measuring the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a variety of tissues. The presence of a CSNA or CSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a CSNA or CSP is associated with colon cancer. Similarly, the presence of a CSNA or

10 CSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a CSNA or CSP is associated with colon cancer. Further, the presence of a structurally altered CSNA or CSP that is associated with colon cancer is also indicative of metastasis.

In general, if high expression relative to a control of a CSNA or CSP is indicative

15 of metastasis, an assay for metastasis is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is

20 considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The CSNA or CSP of this invention may be used as element in an array or a

25 multi-analyte test to recognize expression patterns associated with colon cancers or other colon related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of colon disorders.

30 *Staging*

The invention also provides a method of staging colon cancer in a human patient.

The method comprises identifying a human patient having colon cancer and analyzing

cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more CSNAs or CSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more CSNAs or CSPs is determined for each stage to obtain a standard expression level for each CSNA and CSP. Then, the CSNA or CSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The CSNA or CSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the CSNAs and CSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a CSNA or CSP to determine the stage of a colon cancer.

Monitoring

Further provided is a method of monitoring colon cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the colon cancer. The method comprises identifying a human patient that one wants to monitor for colon cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more CSNAs or CSPs, and comparing the CSNA or CSP levels over time to those CSNA or CSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a CSNA or CSP that are associated with colon cancer.

If increased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a

-102-

cancerous lesion, then detecting an decrease in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of CSNAs or CSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples.

- 5 Monitoring a patient for onset of colon cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a CSNA and/or CSP. The present invention
10 provides a method in which a test sample is obtained from a human patient and one or more CSNAs and/or CSPs are detected. The presence of higher (or lower) CSNA or CSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly colon cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more CSNAs and/or
15 CSPs of the invention can also be monitored by analyzing levels of expression of the CSNAs and/or CSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

20 *Detection of Genetic Lesions or Mutations*

The methods of the present invention can also be used to detect genetic lesions or mutations in a CSG, thereby determining if a human with the genetic lesion is susceptible to developing colon cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing colon cancer. Genetic lesions can be detected,
25 for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the CSGs of this invention, a chromosomal rearrangement of CSG, an aberrant modification of CSG (such as of the methylation pattern of the

Methods of Detecting Noncancerous Colon Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a sample from a patient suspected of having or known to have a noncancerous colon disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP, comparing the expression level or structural alteration of the CSNA or CSP to a normal colon control, and then ascertaining whether the patient has a noncancerous colon disease. In general, if high expression relative to a control of a CSNA or CSP is indicative of a particular noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of a noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a CSNA and/or CSP is associated with a particular noncancerous colon disease by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining which CSNAs and/or CSPs are expressed in the tissue at either a higher or a lower level than in normal colon tissue. In another embodiment, one may determine whether a CSNA or CSP exhibits structural alterations in a particular noncancerous colon disease state by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining the structural alterations in one or more CSNAs and/or CSPs relative to normal colon tissue.

Methods for Identifying Colon Tissue

In another aspect, the invention provides methods for identifying colon tissue. These methods are particularly useful in, *e.g.*, forensic science, colon cell differentiation
5 and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is colon tissue or has colon tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising colon tissue or having colon tissue-like characteristics, determining whether the sample expresses one or more CSNAs
10 and/or CSPs, and, if the sample expresses one or more CSNAs and/or CSPs, concluding that the sample comprises colon tissue. In a preferred embodiment, the CSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 75 through 124, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 74, or a
15 hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a CSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a CSP is expressed. Determining whether a sample
20 expresses a CSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the CSP has an amino acid sequence selected from SEQ ID NO: 75 through 124, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two CSNAs and/or CSPs is determined. In a more preferred embodiment, the
25 expression of at least three, more preferably four and even more preferably five CSNAs and/or CSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is colon tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are
30 recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into colon tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new colon tissue by tissue engineering. These agents include,

e.g., growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

5 Methods for Producing and Modifying Colon Tissue

In another aspect, the invention provides methods for producing engineered colon tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a CSNA or a CSG into the cells, and growing the cells under conditions in
10 which they exhibit one or more properties of colon tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal colon tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered colon tissue or cells comprises one of these cell types. In another embodiment, the engineered colon tissue or cells comprises more than one colon cell
15 type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the colon cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more CSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules
20 encode CSPs having amino acid sequences selected from SEQ ID NO: 75 through 124, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 74, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a CSG is introduced into the cells.

-106-

antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a CSNA or part thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 74, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a CSP or fragment thereof. In a more preferred embodiment, the CSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 75 through 124, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-CSP antibody, preferably an antibody that specifically binds to a CSP having an amino acid that is selected from the group consisting of SEQ ID NO: 75 through 124, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and

-107-

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (PovidoneTM), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

5 Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing,
10 together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

15 For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum
20 albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and
25 administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate,
30 isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of
5 the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature
10 of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered
15 topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include
20 chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from
25 about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be
30 administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for
5 treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the
10 present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an
15 appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for
20 example CSP polypeptide, fusion protein, or fragments thereof, antibodies specific for CSP, agonists, antagonists or inhibitors of CSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present
25 invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the
30 population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic

-111-

effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After
5 administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

10 The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting
15 pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody
20 agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (e.g., 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will
25 employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to
30 the patient. The pharmaceutical compositions of the present invention can be

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of colon function. As used herein, “treating” includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term “treating” encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

10 *Gene Therapy and Vaccines*

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of “naked” nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a CSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a CSP are administered, for example, to complement a deficiency in the native CSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid

molecule encodes a CSP having the amino acid sequence of SEQ ID NO: 75 through 124, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a CSP, fusions, or fragments thereof can
5 be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in CSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a CSP having the amino acid sequence of SEQ ID NO: 75 through 124, or a fragment, fusion protein, allelic variant or homolog thereof.

10 *Antisense Administration*

Antisense nucleic acid compositions, or vectors that drive expression of a CSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a CSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

15 Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a CSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of
20 sequence-specific hybridization to CSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

25 Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the CSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical
30 compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

-114-

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a CSP, preferably a CSP comprising an amino acid sequence of SEQ ID NO: 75 through 124, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule
5 having a nucleotide sequence of SEQ ID NO: 1 through 74, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a CSP, a
10 fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant CSP defect.

Protein compositions are administered, for example, to complement a deficiency in native CSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to CSP. The immune response can
15 be used to modulate activity of CSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate CSP.

In a preferred embodiment, the polypeptide is a CSP comprising an amino acid
20 sequence of SEQ ID NO: 75 through 124, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 74, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

25 Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example,
30 to antagonize activity of CSP or to target them to specific sites of CSP presence

-115-

CSP comprising an amino acid sequence of SEQ ID NO: 75 through 124, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a CSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 74, or a part, allelic
5 variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a CSP or have a modulatory effect on the expression or activity of a CSP. Modulators which decrease the expression or activity of CSP (antagonists) are believed to be useful in treating colon cancer. Such screening assays are known to those of skill in
10 the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a CSP can also be designed, synthesized and tested for use in the imaging and treatment of colon cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the CSPs identified herein. Molecules
15 identified in the library as being capable of binding to a CSP are key candidates for further evaluation for use in the treatment of colon cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a CSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of CSP is
20 administered. Antagonists of CSP can be produced using methods generally known in the art. In particular, purified CSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a CSP.

In other embodiments a pharmaceutical composition comprising an agonist of a
25 CSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP comprising an amino acid sequence of SEQ ID NO: 75 through 124, or a fusion protein, allelic variant, homolog, analog or derivative
30 thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP encoded by a nucleic acid molecule

-116-

having a nucleotide sequence of SEQ ID NO: 1 through 74, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Colon Tissue

The invention also provides a method in which a polypeptide of the invention, or
5 an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the colon or to specific cells in the colon. In a preferred embodiment, an anti-CSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if colon tissue needs to be selectively destroyed. This would be useful for targeting and killing colon cancer cells.
10 In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting colon cell function.

In another embodiment, an anti-CSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring colon function, identifying colon cancer tumors,
15 and identifying noncancerous colon diseases.

EXAMPLES

Example 1: Gene Expression analysis

-117-

	DEX0289_14	DEX0133_9	DEX0289_83
	DEX0289_15	flex DEX0133_9	DEX0289_84
	DEX0289_16	DEX0133_10	DEX0289_85
	DEX0289_17	flex DEX0133_10	
5	DEX0289_18	DEX0133_11	DEX0289_86
	DEX0289_19	flex DEX0133_11	
	DEX0289_20	DEX0133_12	DEX0289_87
	DEX0289_21	flex DEX0133_12	
	DEX0289_22	DEX0133_13	DEX0289_88
10	DEX0289_23	DEX0133_14	DEX0289_89
	DEX0289_24	flex DEX0133_14	
	DEX0289_25	DEX0133_15	DEX0289_90
	DEX0289_26	flex DEX0133_15	DEX0289_91
	DEX0289_27	DEX0133_17	DEX0289_92
15	DEX0289_28	flex DEX0133_17	DEX0289_93
	DEX0289_29	DEX0133_19	DEX0289_94
	DEX0289_30	flex DEX0133_19	
	DEX0289_31	DEX0133_20	DEX0289_95
	DEX0289_32	flex DEX0133_20	
20	DEX0289_33	DEX0133_21	DEX0289_96
	DEX0289_34	flex DEX0133_21	DEX0289_97
	DEX0289_35	DEX0133_22	DEX0289_98
	DEX0289_36	flex DEX0133_22	
	DEX0289_37	DEX0133_24	DEX0289_99
25	DEX0289_38	flex DEX0133_24	
	DEX0289_39	DEX0133_25	DEX0289_100
	DEX0289_40	flex DEX0133_25	
	DEX0289_41	DEX0133_26	DEX0289_101
	DEX0289_42	flex DEX0133_26	
30	DEX0289_43	DEX0133_27	DEX0289_102
	DEX0289_44	flex DEX0133_27	
	DEX0289_45	DEX0133_28	
	DEX0289_46	DEX0133_29	DEX0289_103
	DEX0289_47	flex DEX0133_29	
35	DEX0289_48	DEX0133_30	DEX0289_104
	DEX0289_49	flex DEX0133_30	DEX0289_105
	DEX0289_50	DEX0133_33	DEX0289_106
	DEX0289_51	DEX0133_34	DEX0289_107

-118-

DEX0289_62 flex DEX0133_39
 DEX0289_63 DEX0133_41 DEX0289_116
 DEX0289_64 flex DEX0133_41
 DEX0289_65 DEX0133_42 DEX0289_117
 5 DEX0289_66 flex DEX0133_42 DEX0289_118
 DEX0289_67 DEX0133_43 DEX0289_119
 DEX0289_68 DEX0133_44 DEX0289_120
 DEX0289_69 flex DEX0133_44 DEX0289_121
 DEX0289_70 DEX0133_45 DEX0289_122
 10 DEX0289_71 flex DEX0133_45
 DEX0289_72 DEX0133_46 DEX0289_123
 DEX0289_73 flex DEX0133_46
 DEX0289_74 CLN113 DEX0289_124

15 The expression levels from the Incyte LifeSeq database are listed below:

DEX0289_10 SEQ ID NO: 10 MAM .0019 LIV .0019 KID .0026 PRO .0028
 DEX0289_11 SEQ ID NO: 11 MAM .0019 LIV .0019 KID .0026 PRO .0028
 DEX0289_13 SEQ ID NO: 13 BRN .0002 BRN .0002 CON .0011 CON .0011
 20 DEX0289_14 SEQ ID NO: 14 KID .0103 BON .0169 OVR .0195
 DEX0289_15 SEQ ID NO: 15 KID .0103 BON .0169 OVR .0195
 DEX0289_18 SEQ ID NO: 18 PRO .0006
 DEX0289_19 SEQ ID NO: 19 PRO .0006
 DEX0289_20 SEQ ID NO: 20 NOS .022 INL .0224
 25 DEX0289_22 SEQ ID NO: 22 PAN .0012 BLD .0016 BMR .0064
 DEX0289_25 SEQ ID NO: 25 KID .009 LIV .0151 BON .0169 ESO .0204
 DEX0289_26 SEQ ID NO: 26 KID .009 LIV .0151 BON .0169 ESO .0204
 DEX0289_3 SEQ ID NO: 3 OVR .0041 KID .0051 PRO .0056 THR .0068
 DEX0289_33 SEQ ID NO: 33 LIV .0057 SPL .0063 UNC .008 OVR .0082
 30 DEX0289_34 SEQ ID NO: 34 LIV .0057 SPL .0063 UNC .008 OVR .0082
 DEX0289_35 SEQ ID NO: 35 THR .0091 UTR .0132 TON .0299
 DEX0289_36 SEQ ID NO: 36 THR .0091 UTR .0132 TON .0299
 DEX0289_37 SEQ ID NO: 37 THR .0091 BMR .0129 LMN .0139
 DEX0289_39 SEQ ID NO: 39 INL .0006 GLB .0185
 35 DEX0289_4 SEQ ID NO: 4 ADR .0015
 DEX0289_40 SEQ ID NO: 40 INL .0006 GLB .0185
 DEX0289_43 SEQ ID NO: 43 LMN .0028 UNC .004 LIV .0057 INT .015
 DEX0289_44 SEQ ID NO: 44 LMN .0028 UNC .004 LIV .0057 INT .015
 DEX0289_46 SEQ ID NO: 46 INS .001 INS .001 UTR .0013 BLV .0016
 40 DEX0289_47 SEQ ID NO: 47 INS .001 INS .001 UTR .0013 BLV .0016
 DEX0289_48 SEQ ID NO: 48 INL .0051 LIV .0057
 DEX0289_49 SEQ ID NO: 49 INL .0051 LIV .0057

-119-

	DEX0289_5	SEQ ID NO: 5	ADR .0015				
	DEX0289_51	SEQ ID NO: 51	UNC .008				
	DEX0289_52	SEQ ID NO: 52	UNC .008				
	DEX0289_53	SEQ ID NO: 53	SAG .079	SAG .079	PIT .3246	PIT .3246	
5	DEX0289_54	SEQ ID NO: 54	SAG .079	SAG .079	PIT .3246	PIT .3246	
	DEX0289_55	SEQ ID NO: 55	CRD .0023	TST .0027	INS .0048	CON .0068	
	DEX0289_59	SEQ ID NO: 59	UNC .004	ESO .0051	LIV .0094	SYN .0112	
	DEX0289_60	SEQ ID NO: 60	UNC .004	ESO .0051	LIV .0094	SYN .0112	
	DEX0289_61	SEQ ID NO: 61	LNG .0006				
10	DEX0289_62	SEQ ID NO: 62	LNG .0006				
	DEX0289_68	SEQ ID NO: 68	INS .0789				
	DEX0289_70	SEQ ID NO: 70	OVR .0031				
	DEX0289_71	SEQ ID NO: 71	OVR .0031				
	DEX0289_72	SEQ ID NO: 72	PRO .0017	OVR .0021			
15	DEX0289_74	SEQ ID NO: 74	FTS .0003	CON .0011	LIV .0019	OVR .0021	
	DEX0289_8	SEQ ID NO: 8	BRN .0004	PRO .0006	CON .0011	LIV .0019	
	DEX0289_9	SEQ ID NO: 9	BRN .0004	PRO .0006	CON .0011	LIV .0019	

Abbreviation for tissues:

- 20 BLO Blood; BRN Brain; CON Connective Tissue; CRD Heart; FTS Fetus; INL Intestine, Large; INS Intestine, Small; KID Kidney; LIV Liver; LNG Lung; MAM Breast; MSL Muscles; NRV Nervous Tissue; OVR Ovary; PRO Prostate; STO Stomach; THR Thyroid Gland; TNS Tonsil / Adenoids; UTR Uterus

25 Example 2: Relative Quantitation of Gene Expression

- Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity
- 30 of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate
- 35 dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this

-120-

endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence
5 Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain
10 reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of
15 expression of the CSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal tissue (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the CSNA in pairs of matching samples and 1
20 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal tissue (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, CSNAs show a high degree of tissue
25 specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in
30 the cancer sample compared to the normal adjacent).

-121-

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 74 being diagnostic markers for cancer.

Cln113; DEX0289_74; DEX0289_124; LifeSeq Gold Gene ID# 1394652

5 X73079 g456345 Human encoding Polymeric Ig receptor "Cln113"

DNA sequence for Cln113

Sequence available from GenBank database

Homo sapiens encoding Polymeric immunoglobulin receptor. Ac# X73079

DEX0289_74

10 Experiments on Cln113 are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. Routine molecular biology techniques of the following example can be carried out as described in standard laboratory manuals, such as Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press,
15 Cold Spring Harbor, N.Y. (1989).

Relative Quantitation of Gene Expression.

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye
20 and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA).

We use amplification of an endogenous control to standardize the amount of sample
25 RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. We either use cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) or 18S ribosomal RNA (rRNA) as this endogenous control. To calculate relative quantitation between all the samples studied, we used the target RNA levels for one sample as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained

20 with the standard curve method or the comparative method (see Bulletin #2: ABI PRISM

We evaluated the tissue distribution, and the level of the target gene for every example in normal and cancer tissue. Total RNA was extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA was prepared with reverse transcriptase and the polymerase chain
 5 reaction was done using primers and Taqman probe specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

Table 1. The absolute numbers are relative levels of expression of Cln113 in 12 normal
 10 different tissues. All the values are compared to normal colon (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

Tissue	NORMAL
Colon-Ascending	1.00
Endometrium	0.00
Kidney	0.07
Liver	0.00
Ovary	0.00
Pancreas	0.00
Prostate	0.00
Small Intestine	0.28
Spleen	0.00
Stomach	0.30
Testis	0.00
Uterus	0.00

15 The relative levels of expression in Table 1 show that overall gene expression levels of Cln113 in the RNA samples from different normal tissues are low. All the normal tissues analyzed had lower levels of expression compared to normal colon which was used as a calibrator with a relative expression level of 1. These results demonstrated that Cln113 mRNA expression is highly specific for colon.

20 The absolute numbers in Table 1 were obtained analyzing pools of samples of a particular tissue from different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in Table 2.

Table 2. The absolute numbers are relative levels of expression of Cln113 in 48 pairs of matching samples. All the values are compared to normal colon (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

5

Sample ID	Cancer Type	Tissue	CANCER	MATCHING NORMAL ADJACENT
Stomach	StoMT54	Stomach-1	0.12	0.00
Small Intestine	SmInt21XA	Small Intestine -1	0.05	0.00
Colon- Ascending (A)	ClnAS45	Colon-1	0.00	0.23
Colon-Cecum (B)	ClnCM67	Colon-2	0.29	0.00
Colon-Cecum (C)	ClnB56	Colon-3	0.8	0.00
Colon- Ascending (B)	ClnAS67	Colon-4	0.7	0.3
Colon- Ascending (B)	ClnAS12	Colon-5	0.19	0.55
Colon- Ascending (C)	ClnAS43	Colon-6	4.71	0.00
Colon- Ascending (C)	ClnAS46	Colon-7	0.00	3.06
Colon- Ascending (D)	ClnAS89	Colon-8	1.1	0.00
Colon- Ascending (D)	ClnAC19	Colon-9	0.00	0.99
Colon- Transverse (B)	ClnTX01	Colon-10	0.08	0.73
Colon- Transverse (B)	ClnTX89	Colon-11	1.3	0.00
Colon- Transverse (C)	ClnTX67	Colon-12	0.00	0.4
Colon- Descending (C)	ClnDC63	Colon-13	1.02	0.00
Colon-Sigmoid (C)	ClnSG27	Colon-14	0.35	0.41
Colon-Sigmoid (B)	ClnSG20	Colon-15	1.59	1.33
Colon-Sigmoid (D)	ClnSG45	Colon-16	0.00	0.9
Colon- Rectosigmoid (A)	ClnB34	Colon-17	0.1	0.00
Colon-Rectum (A)	ClnCXGA	Colon-18	1.97	0.37
Colon-Rectum (B)	ClnRC67	Colon-19	0.00	0.00
Colon- Rectosigmoid (C)	ClnC9XR	Colon-20	0.00	0.00

-124-

(C)				
Colon-Rectum (D)	ClnRC89	Colon-22	0.00	0.4
Bladder	Bld32XK	Bladder-1	0.00	0.00
Bladder	Bld46XK	Bladder-2	0.00	0.00
Cervix	CvxKS83	Cervix-1	0.00	0.00
Cervix	CvxKS52	Cervix-2	0.00	0.00
Endometrium	endo10479	Endometrium-1	0.00	0.00
Endometrium	endo12XA	Endometrium-2	0.00	0.0
Kidney	Kid11XD	Kidney-1	0.00	0.00
Kidney	Kid10XD	Kidney-2	0.00	0.00
Kidney	Kid107XD	Kidney-3	0.00	0.01
Kidney	Kid109XD	Kidney-4	0.19	0.11
Kidney	Kid106XD	Kidney-5	0.05	0.02
Liver	Liv42X	Liver-1	0.00	0.00
Liver	Liv15XA	Liver-2	0.00	0.00
Liver	Liv94XA	Liver-3	0.00	0.00
Lung	LngAC11	Lung-1	0.00	0.00
Lung	Lng90X	Lung-2	0.00	0.00
Lung	Lng60XL	Lung-3	0.00	0.00
Lung	LNG47XQ	Lung-4	0.00	0.00
Mammary Gland	Mam12X	Mammary gland-1	0.00	0.00
Mammary Gland	Mam14DN	Mammary gland-2	0.00	0.00
Prostate	Pro12B	Prostate-1	0.00	0.00
Testis	Tst39X	Testis-1	0.00	0.00
Uterus	Utr85XU	Uterus-1	0.08	0.00
Uterus	Utr135XO	Uterus-2	0.00	0.00

0= Negative

In the analysis of matching samples, the higher levels of expression were in colon showing a high degree of tissue specificity for colon tissue. These results confirm the tissue specificity results obtained with normal pooled samples (Table 1).

Furthermore, we compared the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 2 shows overexpression of Cln113 in 10 of the colon cancer tissues compared with their respective normal adjacent (colon samples # 2, 3, 4, 6, 8, 11, 13, 17, 18, and 21). There is overexpression in the cancer tissue for 45% of the colon matching samples tested (total of 22 colon matching samples).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in 45% of the colon adenocarcinoma matching samples tested are believed to make Cln113 a good diagnostic marker for colon cancer.

- 5 forward: hitting DEX0289_74 SEQ ID NO. 74 (bp1441-1462)
reverse: hitting DEX0289_74 SEQ ID NO. 74 (bp1603-1585)

Example 3: Protein Expression

- 10 The CSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the CSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the CSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH₂-terminus of the coding sequence of CSNA, and six histidines, flanking the COOH-terminus of the coding sequence of CSNA, are incorporated to serve as initiating
15 Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

- 20 Large-scale purification of CSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column volumes of wash buffer. CSP was eluted stepwise with various concentration imidazole buffers.

25 Example 4: Protein Fusions

- Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4
30 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without

a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.,* WO 5 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to 10 culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; 15 however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to 20 identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, 25 protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be 30 used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and

-127-

Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

Based on the underlying sequences found by mRNA subtractions the following extended nucleotide sequences and predicted amino acid sequences were determined.

- 5 The chromosomal locations were determined for several of the sequences. Specifically:

- DEX0289_3 chromosome 6
- DEX0289_6 chromosome 7
- DEX0289_10 chromosome 13
- 10 DEX0289_16 chromosome 9
- DEX0289_17 chromosome 9
- DEX0289_20 chromosome 3
- DEX0289_22 chromosome 4
- DEX0289_23 chromosome 6
- 15 DEX0289_26 chromosome 3
- DEX0289_29 chromosome 9
- DEX0289_30 chromosome 9
- DEX0289_31 chromosome 7
- DEX0289_32 chromosome 7
- 20 DEX0289_34 chromosome 13
- DEX0289_43 chromosome 1
- DEX0289_44 chromosome 1
- DEX0289_45 chromosome 3
- DEX0289_49 chromosome 9
- 25 DEX0289_51 chromosome 7
- DEX0289_52 chromosome 7
- DEX0289_55 chromosome 3
- DEX0289_56 chromosome 3
- DEX0289_57 chromosome 9
- 30 DEX0289_58 chromosome 9
- DEX0289_63 chromosome 7
- DEX0289_64 chromosome 7
- DEX0289_66 chromosome 4
- DEX0289_67 chromosome 3
- 35 DEX0289_68 chromosome 16
- DEX0289_69 chromosome 16
- DEX0289_72 chromosome 4
- DEX0289_74 chromosome 1

- 40 The predicted antigenicity for the amino acid sequences is as follows:

-128-

	98-108	1.05	11
	DEX0289_85 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
5	29-44	1.14	16
	DEX0289_91 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
10	66-82	1.00	17
	DEX0289_98 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	22-35	1.04	14
15	DEX0289_101 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	5-14	1.04	10
	DEX0289_108 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
20	618-627	1.10	10
	576-611	1.10	36
	330-341	1.07	12
	488-498	1.06	11
25	DEX0289_113 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	47-56	1.15	10
30	DEX0289_115 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	16-30	1.03	15
	DEX0289_120 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
35	12-24	1.10	13
	DEX0289_121 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
40	54-67	1.27	14
	DEX0289_124 Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
45	372-382	1.30	11
	99-110	1.26	12
	42-63	1.15	22
	506-516	1.15	11
	270-279	1.12	10
	385-394	1.12	10
50	484-504	1.03	21
	179-193	1.00	15

The predicted helicity for the amino acid sequences is listed below:

	DEX0289_80	PredHel=1	Topology=i9-31o
55	DEX0289_88	PredHel=1	Topology=o15-32i
	DEX0289_108	PredHel=9	Topology=i35-57o63-85i92-109o170-192i199-216o226-248i261-
	283o298-320i359-378o		
	DEX0289_112	PredHel=1	Topology=i2-19o

DEX0289_124 PredHel=1 Topology=o639-661i

- Examples of post-translational modifications (PTMs) of the BSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the LSPs of the invention (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html most recently accessed October 23, 2001). For full definitions of the PTMs see <http://www.expasy.org/cgi-bin/prosite-list.pl> most recently accessed October 23, 2001.
- DEX0289_100 Myristyl 13-18;36-41; Pkc_Phospho_Site 26-28;
 - DEX0289_101 Ck2_Phospho_Site 46-49; Myristyl 42-47; Pkc_Phospho_Site 20-22;46-48;58-60;
 - DEX0289_102 Amidation 29-32;
 - DEX0289_103 Ck2_Phospho_Site 25-28; Myristyl 33-38;
 - 15 DEX0289_104 Asn_Glycosylation 35-38; Myristyl 39-44; Pkc_Phospho_Site 42-44;
 - DEX0289_105 Asn_Glycosylation 18-21; Ck2_Phospho_Site 76-79; Myristyl 19-24;39-44;
 - Pkc_Phospho_Site 4-6;
 - DEX0289_106 Ck2_Phospho_Site 27-30;
 - DEX0289_107 Myristyl 5-10;11-16;
 - 20 DEX0289_108 Asn_Glycosylation 111-114;533-536;598-601; Camp_Phospho_Site 263-266;620-623;
 - Ck2_Phospho_Site 18-21;138-141;155-158;215-218;247-250;524-527;631-634;684-687;
 - Myristyl 245-250;380-385;404-409;409-414;446-451;464-469;534-539;632-637;
 - Pkc_Phospho_Site 4-6;32-34;113-115;122-124;219-221;249-251;332-334;519-521;606-608;611-613;619-621;
 - 25 DEX0289_109 Asn_Glycosylation 20-23; Pkc_Phospho_Site 2-4;11-13;22-24;34-36;
 - DEX0289_111 Asn_Glycosylation 105-108; Camp_Phospho_Site 48-51; Ck2_Phospho_Site 13-16;
 - Glycosaminoglycan 93-96; Myristyl 108-113; Pkc_Phospho_Site 87-89;
 - DEX0289_113 Ck2_Phospho_Site 97-100; Myristyl 20-25; Pkc_Phospho_Site 48-50;67-69;97-99;
 - DEX0289_114 Ck2_Phospho_Site 6-9; Pkc_Phospho_Site 6-8;
 - 30 DEX0289_115 Ck2_Phospho_Site 29-32;
 - DEX0289_116 Camp_Phospho_Site 11-14;15-18; Ck2_Phospho_Site 14-17;19-22; Pkc_Phospho_Site 14-16;
 - DEX0289_117 Ck2_Phospho_Site 30-33; Pkc_Phospho_Site 35-37;
 - DEX0289_118 Pkc_Phospho_Site 15-17;
 - 35 DEX0289_120 Pkc_Phospho_Site 27-29;
 - DEX0289_121 Ck2_Phospho_Site 55-58; Myristyl 77-82;
 - DEX0289_122 Camp_Phospho_Site 19-22;
 - DEX0289_123 Asn_Glycosylation 11-14;
 - DEX0289_124 Amidation 286-289; Asn_Glycosylation 83-86;90-93;135-138;186-189;421-424;469-472;499-502; Camp_Phospho_Site 161-164;375-378;732-735; Ck2_Phospho_Site 101-104;274-277;414-417;432-435;443-446;451-454;489-492;508-511;535-538;629-632;682-685;684-687;721-724;734-737;735-738; Myristyl 33-38;59-64;62-67;111-116;216-221;361-366;364-369;389-394;468-473;545-550;620-625;635-640;648-653;659-664;693-698; Pkc_Phospho_Site 17-19;51-53;75-77;105-107;188-190;228-230;380-382;432-434;477-479;636-638;673-675;676-678;734-736; Tyr_Phospho_Site 211-218;239-246;316-323;536-542;736-743;
 - 45 DEX0289_75 Asn_Glycosylation 4-7; Pkc_Phospho_Site 18-20;
 - DEX0289_76 Pkc_Phospho_Site 24-26;
 - DEX0289_77 Camp_Phospho_Site 5-8;
 - 50 DEX0289_78 Ck2_Phospho_Site 138-141; Myristyl 26-31;
 - DEX0289_79 Ck2_Phospho_Site 30-33; Pkc_Phospho_Site 5-7;53-55;

-130-

- DEX0289_81 Pkc_Phospho_Site 14-16;
 DEX0289_82 Asn_Glycosylation 11-14; Ck2_Phospho_Site 13-16;
 DEX0289_83 Myristyl 13-18; Pkc_Phospho_Site 18-20;
 5 DEX0289_84 Asn_Glycosylation 13-16;73-76;166-169; Ck2_Phospho_Site 18-21; Pkc_Phospho_Site 132-134;
 DEX0289_85 Camp_Phospho_Site 23-26; Myristyl 33-38;39-44; Pkc_Phospho_Site 21-23;34-36;
 DEX0289_86 Asn_Glycosylation 16-19; Ck2_Phospho_Site 18-21; Myristyl 5-10;
 DEX0289_87 Asn_Glycosylation 23-26; Pkc_Phospho_Site 20-22;25-27;
 DEX0289_88 Asn_Glycosylation 10-13; Pkc_Phospho_Site 7-9;
 10 DEX0289_91 Ck2_Phospho_Site 57-60; Myristyl 69-74; Pkc_Phospho_Site 5-7;73-75;
 DEX0289_92 Myristyl 2-7;
 DEX0289_93 Amidation 65-68; Camp_Phospho_Site 37-40;72-75; Ck2_Phospho_Site 26-29;46-49;75-78; Pkc_Phospho_Site 25-27;65-67;75-77;
 DEX0289_94 Asn_Glycosylation 38-41; Ck2_Phospho_Site 5-8; Pkc_Phospho_Site 21-23;
 15 DEX0289_95 Ck2_Phospho_Site 15-18;
 DEX0289_96 Pkc_Phospho_Site 31-33;
 DEX0289_97 Myristyl 20-25; Pkc_Phospho_Site 13-15;40-42;
 DEX0289_98 Myristyl 16-21; Pkc_Phospho_Site 33-35.

Example 6: Method of Determining Alterations in a Gene Corresponding to a

20 Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1
 25 through 74. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4
 30 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156
 35 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99

(1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping
5 are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the
10 genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

15 Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 $\mu\text{g/ml}$. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The
20 coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature.
25 The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is
30 prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear

scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion
5 consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

10 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 $\mu\text{g/kg/day}$ to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given
15 continuously, the secreted polypeptide is typically administered at a dose rate of about 1 $\mu\text{g/kg/hour}$ to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the
20 desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a
25 non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release
30 systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers

of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-
5 release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and
10 EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is
15 formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and
20 other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such
25 carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at
30 the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine

or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

- 5 For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 $\mu\text{g/kg}$ of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

- 10 Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

- For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0
15 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

- One method of gene therapy transplants fibroblasts, which are capable of
20 expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room
25 temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

- At this time, fresh media is added and subsequently changed every several days.
30 After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine

-136-

sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified
5 using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions
10 appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf
15 serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the
20 media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

25 If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or

30 A having been grown to confluence on cytotoxic 2 microcarrier beads

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences
5 into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622;
10 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers
15 injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell,
20 including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

25 The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target
30 cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

-138-

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue.

5 Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to
10 the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin
15 fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 $\mu\text{g/kg}$ body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and
20 more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the
25 parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

30 The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard

recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by
5 intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin
10 is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different
15 mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

20 **Example 13: Transgenic Animals**

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific
25 embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection
30 (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene

transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a
5 gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993)); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989)); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its
10 entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

15 The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a
20 particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous
25 gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous

-141-

required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512

-142-

(1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered

-143-

cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; 5 genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible 10 cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

15 Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

20 All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, 25 which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

-144-

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes
5 an amino acid sequence of SEQ ID NO: 75 through 124;
 - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID
NO: 1 through 74;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid
molecule of (a) or (b); or
10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is a cDNA.
15
3. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of a colon specific nucleic acid
(CSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule according to claim 1
under conditions in which the nucleic acid molecule will selectively hybridize to a colon
specific nucleic acid; and
 - 30 (b) detecting hybridization of the nucleic acid molecule to a CSNA in the
sample, wherein the detection of the hybridization indicates the presence of a CSNA in
the sample

-145-

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

5

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60%

15 sequence identity to of SEQ ID NO: 75 through 124; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic

acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 74

-146-

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the colon specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of colon cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with colon cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein

17. A method of treating a patient with colon cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein

SEQUENCE LISTING

<110> Macina, Roberto
 Recipon, Herve
 Pluta, Jason
 Ghosh, Malavika
 Sun, Yongming
 Liu, Chenghua
 diaDexus, Inc.

<120> Compositions and Methods Relating to Colon Specific Genes and
 Proteins

<130> DEX-0290

<150> 60/252,505

<151> 2000-11-22

<160> 124

<170> PatentIn version 3.1

<210> 1

<211> 421

<212> DNA

<213> Homo sapien

<400> 1

```

cgtggtcgcg ggccagaggt accttcctcc aatgttggtt tcagcccaca ccattactag      60
atgatcgccct aggctcttct gaagctctct ctaaactcat aattattggt tggaccctgg      120
catgttaact aaacttaatt gtgccaagtg atgggaaatg aaactgtaca gttttatgtg      180
gcaacgaatg gtaatccccg caaaacagaa tgacagatac agtgatgggt aagtagatgt      240
tactgccctg ttaattgggt ccgaagcata agatacacct gaaaaataat gtgaaaactg      300
aatttgcctt tgatttgaaa aatctagaga atcagcatat aatgtttggt aatgttctta      360
agctggtaaa tatcattaag agaaatggac acatataaga taagtttgtg tgcataattg      420
t                                                                                   421

```

<210> 2

<211> 612

<212> DNA

<213> Homo sapien

<400> 2

```

acattttaat ttacatgtgt gtagaacata gatgagaact ctgggaaaac ttgggaatgg      60
caaccaacca aaatcatttt taatcattta ttagaaattt ctcaatattg tgtctttttc      120
ttttgaaact ctaaactt cagaaaaaaa cactatcagt gtagttcatg ttagtataat      180
tatagattta catatatttg aatagttaat ttgctttggt ttacacgtag cccactgcct      240

```


cattataggt aaaaggcatt tataactgct caggggatta cgagaactca actgaaactg 300
 aatTTTTgta acaagaatgt taatagtggc aaagtcctct gtcagtaaac tctttaagct 360
 tggTgcgcga aagagtcttt aaatgggggc tgatttcaag taacctaaaa gactgtgtta 420
 tccgaagaag aagggtcccc aaattggagt aagaatggga gaaaaaaaaa aagtgtctatt 480
 tccctggcga gttgggggga attgcccccc tacagagttt gtatcactga attagctgct 540
 tttgtttctt tttttttggg caggggtttg ggaggggggt tggttggtca actggttttc 600
 caaaacgtgc tc 612

<210> 3
 <211> 1100
 <212> DNA
 <213> Homo sapien

<400> 3
 gataaaaccg caacaaaaac atgtaagaaa taaaatagaa atgctttata tatttttagtt 60
 taaatttatg tatcacctca ttgtgactta tttttccat tataccatta gtcagatttg 120
 aataacgagg ttttgaaagg ataaaacctt ttctccaatg acaggattat ataattgcta 180
 ttggcaatgt agcctgggtgc ttcattgagac ctatgctaaa tgttactgga gagttcttga 240
 agccagggat accatatcag gaactattca ggatctatga tattttctga ggtaactggg 300
 taatagaata tcaaattgct gctatctcgg acctattgtt aaaggatgat gctttgccta 360
 tgtaatagga tatatcctaa gtggggatgt gtatatttca ggaacttta ttcacaagta 420
 tatattgata tctgatgtgt gtatagtaca tctgttggtt atgtacattt taatttacat 480
 gttgtgtaga acatagatga gaactctggg aaaacttggg aatggcaacc aaccaaaatc 540
 atttttaatc atttattaga aatttctcaa tattgtgtct ttttcttttg aaactctaaa 600
 cacttcagaa aaaaacacta tcagtgtagt tcatgttagt ataattatag atttacatat 660
 atttgaatag ttaatttgct ttgttttaca cgtagccac tgcctcatta taggtaaaag 720
 gcatttataa ctgctcaggg gattacgaga actcaactga aactgaattt ttgtaacaag 780
 aatgttaata gtggcaaagt cctctgtcag taaactcttt aagcttggtg ccgcaaagag 840
 tctttaaatg ggggctgatt tcaagtaacc taaaagactg tgttatcaga ggaagaggtc 900
 ccaaatttgg agtaaagatg ggagaaaata aatatgtgct atttccttgg cgagttgggg 960
 gaatttgcca ccttacagag tttgtatcac tgaattagct gcttttggtt tttttttttt 1020
 ttttttttgg ccagggtc tagaagcggg ggtttgtgag cgccaccgtg ttttcacaat 1080
 attggtttta atttttttta 1100

<210> 4
 <211> 627
 <212> DNA
 <213> Homo sapien

<400> 4
 acttcgcaat tcataaaaaat aggttttcaa taaatttgaa catacatact cactgaaaaa 60
 agatacttttg taaaaatggc tataaaaaata tgggttaatgg tgggttaact attggattct 120
 gatataatttc atacctatga tctcattttg tttctagttt tactgatata accaaccttg 180
 gacacccaaa gatgttggtt ttattttctga aattactcag ctatagtaaa gtatcaagaa 240
 tagataattta tatttaagaa gactcaccca tcccagacac tgaactcact aattagccgg 300
 tcagaaagat cactaaggaa caatttacia tgcaataaaa gtgatacgct ttactttctg 360
 agtaacagca gagcaagagg ttccataaga atcctggcaa agcaatcttt ccactttcaa 420
 tgttgatcac ttagatcttg tgaaattcgc ggcgatattt agtataaatg actaggaaag 480
 ctattatttg tgcataagag aaacctaact taattatata cataactcaa caatttgctc 540
 agtgcttttt tgtgcattgg gaaattatgt ttccagaaac ccaaacaaaa caaacagtc 600
 gttgaaattt tctttattag actcagt 627

<210> 5
 <211> 1865
 <212> DNA
 <213> Homo sapien

<400> 5
 gaaacttcaa actaatgatt aaatagtaga gggctgctga tcccttctta tatactgcaa 60
 gaataacact taataaagga tgaagaaaga tttgtactga gtctaataaa gaaaatttca 120
 acgactgggt ttgttttggg ttggttttct gaaacataat ttcccaatgc aaaaaaagc 180
 actgagcaaa ttgttgagtt atggatataa ttaagttagg tttctcttat gcacaaataa 240
 tagcttttct agtcatttat actaaaaatc accacgaatt tcacaagatc taagtgatca 300
 acattgaaag tggaaagatt gctttgccag gattottatg gaacctcttg ctctgctgtt 360
 actcagaaag taaagcgtat cacttttatt gcattgtaaa ttgttcctta gtgatctttc 420
 tgaccggcta attagttagt tcagtgtctg ggatgggtga gtcttcttaa atataaatat 480
 ctattcttga tactttacta tagctgagta atttcagaaa taaaaacaac atctttgggt 540
 gtccaagggt gggttatatca gtaaaactag aaacaaaatg agatcatagg tatgaaatat 600
 atcagaatcc aatattaacc caacattaac catattttta tagccatttt tacaaagtat 660
 cttttttcag tgagtatgta tgttcaaatt tattgaaaac ctatttttat gaattgcgaa 720

```

gtacaccaaa tatggcatta atagaactac agccttaact acatgcttat tgtcaggcct      780
ctgagcccaa gctaaaccat cataatcccc tgtgacctgc atgtatacat ccagatggcc      840
tgaagcaagt gaagaattac aaaagaagtg gaaacggccg gttcctgcct taactgatga      900
cattgcgcca ttgtgatttg tttccccacc ttaactgagc gattaacctt gtgaaattcc      960
ttctcctggc tcagaacctc cccactgag caccttggga cccccacccc taccgcaag     1020
agaacaacct cctttgactg taattttcca ctaccaccc aaatcctgta aaacagcccc     1080
accctatct ccttctctg actctctttt tggactcagt cgcctgcac cctggtgaaa     1140
taaacagctt tattgctcac acaaagcctg tttggtggtc tcttcacacg gatgcgagtg     1200
aaatttggtg ccatgactcg gatcggggga cctcccttgg gagatcaatc ccctgtcctc     1260
ctgctctttg ctccgtgaga aagatccacc tacgaccaca ggctctcaga ccaaccagcc     1320
caagaaacat ctaccaatt tcaaactctga cagctttaga gactgcccc accctagctc     1380
tccctgactc atccaacct ttttcattac acacagctga agtgcagggc tgtgcagttg     1440
gaattcttac acaaggacca ggatcgctc ctgtagcctt tttgtccaag caccttgacc     1500
ttactgtttt aggetggtca tcatgtctcc gtgcagcggc ttctgcgccc ctaatacttt     1560
tagaggccct taaaatcaca aactatgctc aactcactct ctacagctct cataatttcc     1620
aaaatctatt ttcttcctca cacctgatgc atgtactttc tgetccctgg ctccctcagc     1680
tgtactcact ctttggtgag tctcccacaa ttaccattat tcctggccgg gacttcaatc     1740
cgcatccca cattattcct gataccacac ctgacctca tgactgcac tctctgatcc     1800
acctgacgtt caccacattt ccccatattt ccttctttcc tgttcctcac cctgatcaca     1860
cttag                                             1865

```

```

<210> 6
<211> 441
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (229)..(230)
<223> a, c, g or t

```

```

<400> 6
acaggagagt gggctctagc aggtggagat acactacgcc ttgacacact tatagaatgg      60
tggagagaaa agaatgggtc cttttgttcc cggttatta tcgtattaga cagcgaaaaat     120
tcaaccctt gggtgaaaga agtgcggaata attaatgacc agtatattgc agtgccaagg     180

```

agcagagttg actaacaac aggtagcata ctctgcaacg caatgcctnn gacccgccac 240
 agctaggtga ctttacaaaa gactgggtag aatataactg caactccagt aataacatct 300
 gctggactga acagggacgc acagtgaaag cagtatatgg tgtgtcaaaa cgggtggagtg 360
 actacactct gcatttgcca acgggaagcg atgtggccaa gcaactggatg ttacactttc 420
 ctctgtattac atatcccta g 441

<210> 7
 <211> 760
 <212> DNA
 <213> Homo sapien

<400> 7
 actggagagt tgttcacaca gatgtttaga cctttctctc tctctctctc tctctttct 60
 tctttctcaa caactctttc acagaggcag tcattttgaa aggttgaaat attgtggctt 120
 taacaaagag cttttttttt ccttaagcaa aatcctttca gaaagaaaca aaatggggaa 180
 gggcagatta agaaatgcat attgtcccaa atccaattct tattaggagg ttaatcatat 240
 ttcaattgag ttaaaattga tgggaagaaa ttcttttagg gtaattcttt ggggattaag 300
 ggatcctggg aagttcctct cagggtaaag gaaaggttta aaagaagatt tgtaatatat 360
 gtctggagag ctatttataa gaaatttaag aggattgttt tgttttcctt ttattaaaga 420
 ttttagcctt tttactttgc aaaaagaaaa ctacaaaagt tttatagata taactttgct 480
 taattgtttg tagaactgtt gtctggaaac gattagctgt agccaaatta tgtggttacg 540
 ttttgcatac ttagaatttg aaaatgcaat atgtgtggtta aatctactgt ttgaaattta 600
 taatggtctc tgatatgatt cgaattttgg taacttttga aagttatttt ccccttttag 660
 tcatggattt ctatttggtt tttaatgtta atttttctag aaagcatctg aattgactag 720
 gcttttccta tataaaaaac tcaaaacttg ttaactctgt 760

<210> 8
 <211> 320
 <212> DNA
 <213> Homo sapien

<400> 8
 cttttttatc tcaaagtcac atacttgtcc atttgtgaca gctgaatacc agaagaatgc 60
 atgtgttget gactagattg ttgatattac aggagctatt gtttgttact ttatttttag 120
 gtgtgatgat ggttttggtt tttatgttta aatgagcctt gtcttttggg gatacatact 180
 gaaatattta tagatgaaat gatctgatgt ctggggaggt ttgcttttaa gtaatagagg 240

agtggggagt agacaggggt atagatgaat caaggttggc catgagttgg taattgttga 300
aactggtgat aggtacctgc 320

<210> 9
<211> 1594
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (538)..(599)
<223> a, c, g or t

<400> 9
caaagatttt tttatgaaac acccgtgttt atgtgcctgg gctgggctct gtatgaaaca 60
ggtaaagctg accccgctca ctactgccc tctaggattt tgttctagga aacttgctag 120
agcctggttc caaaagtaaa caagattgta ttttcatttt tttcttagaa ctatgttatg 180
gacattcagc tcccacatat tctttcacct cttaggcctt gctcaatgaa aataacttgt 240
aaaaaacttg caaaaaactt gctgaaggaa ctgagtgtgt ttagcttggc aacacaaaat 300
tgtggggaac caatgacatc tctcctcaa tatgtgcaa gctgtccctt ggcaaagtag 360
ggcacttatt ctatatgcct tgaaaggaca gaaataggat tattgggtgg aaatgccaag 420
aaggcagact tgagtctgtc tttgtaaaga ctcaagaact ttgtagtagt gtacagttac 480
gagcgtgggc tttggatagt actgggttca aatgcagccg ttgcctcact gcctgacnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
ctacctggta aggcattgtg aggatcaa atgaagcgat acatggctga agcacttaga 660
atgtacttgg catataaata cttggttctc aataattgag aaccagtaat gataatcttt 720
acaataatta gtaacagtca ctatttattg agtgtttaat tatgtgccag aactgaact 780
aaataatttt catatatata gtttatgtaa actactaatt tctgttaata atgacaaata 840
gaattgtcca aaattgaaat tgggtgcttc taaaatagtg aatttttttc tggagagtct 900
gcaagcaaaa attaggtgag cacttgtcag gggaggatgt agttgggggt tcatgcatca 960
ggtgggcaat tggaagagat acgtcctcta aagtcttatt gattctaaga ttttctgggt 1020
ctggagctca ttgataagcg taaggctagt tggagctttt atagtcttta ttgatagcag 1080
tcatccccc cacaccctg atagtaatac actttactat ctgtagtcac gaatgagaaa 1140
gaatttgttt taaagcaaca agggggagaa ttgtgatatt ttaaagcac taacattttt 1200
cttttttatt tcaaagtcac atacttgtca tttgtgaagc tgaataccag aagaatgcat 1260

gtgttgctga ctagattggt gatattaagg agctattggt tgttacttta ttttttaggtg 1320
 tgatgatggt tttgttttta tgtttaaatg agccttgctt tttggagata catactgaaa 1380
 tatttataga tgaaatgac tgatgtctgg ggagggttgc tttaaagtaa tagaggagtg 1440
 gggagtagac aggggtatag atgaatcaag gttggccatg agttggtaat tgttgaaact 1500
 ggtgataggt acatgtgggt ttatatacta ttctgcttct atttatgttt tgaattctcc 1560
 aaataaaact taaaaaagaa gtagaaaaaa aaaa 1594

<210> 10
 <211> 350
 <212> DNA
 <213> Homo sapien

<400> 10
 ccgtaatctg gcaacatccg gggcttacct tcagctctcg cactgtgcgt gagatcgggt 60
 gaggcagtta taagtgagag catgctggac accttgactt tgcagtgcg tggaacagaa 120
 aaagcattca cctcatcatt gaaagagttg gagccgagaa taaaaggtag ttagaaggct 180
 agtgggaagg ggagccggag gccaggaata tagcaactaa caggccctag acagcgatcc 240
 cggcggacag gagaggagga agaactggc actcgggggc caggcgacaa agtcgggggtg 300
 agcacactcc cgatataggc acctccactc tcaaagggcg acagcgagca 350

<210> 11
 <211> 2718
 <212> DNA
 <213> Homo sapien

<400> 11
 agccactgaa ttcccttgcg gccgaggaat tttttttttt tttttttttt tttttgcttc 60
 acaaatgtca attttattga cactagtgc caactaaata caataattgc aaaggaagtg 120
 gaacgtgtca aacagaaatg gtgacaatga gttagaactg cagttgtttc aagggtactac 180
 actattattt aaaaaaaaaa atcacaaaaa gaaaaatgtt atcactacaa gtaggaatta 240
 gaagagagaa attctggcag tctgtctaga ggttaaaaca tttcatgcat ttgtgagttg 300
 ctgttgagaa gttgtttttt atttgtccac cgtaatctgg caacatccgg ggcttacctt 360
 cagctctcgc actgtgcgtg agatcgggtg aggcagttat aagtgcagac atgctggaca 420
 ccttgacttt gcagtgcgtt ggaacagaaa aagcattcac ctcattcatt aaagagttgg 480
 agccgagaat aaaaggtagt tagaaggcta gtgggaaggg gagcggaggc aaggaaatag 540
 caactaacag gccctagaca gcatccggca acagagagga aaagaactgc cactcggggc 600
 aagggaaaaa gtagggggag cacactccga tacagccacc tccactctca aaggccaaca 660

gcgagcacc c ttgctgca ct gcacctggga acacacattt aggggacaga gcagttggaa 720
 gaaatgaggt aacagactat ggttccataa gagagcctgc ctcgccaaga aggcgtgcc 780
 cggttcagaa caatccccac tgtgctacag aggagacagg actcagaaaa cagagggccg 840
 agtgggaact tcagggtcac ctgtgtacct aaacgaagga acagctcagg attagcccac 900
 aggctgctgg gggcaggctt gctgcatttc actcacggag cctaaagatg tcagttaaca 960
 actacttaat atgtgcgctc tgcagacttg gaacgacaaa attaggggtg tcagttggcc 1020
 ttttcccaag acgctactcc agctttgctt acagggccta agaaagaaag ggcaatgggt 1080
 gtgtttaaac agcaagacca agaagccaat aaatatcaaa gtctggtcta gaaatctatc 1140
 agcattttaa ggaagggaaa ggctgaaac tctacagttc agttttgcta atttgagctg 1200
 catctgtgga gaagaggccc cttctctcct tgcaagataa acaatccgag gctttgaaaa 1260
 tgtacaggtg acgtggtcca aacaaaatat gtaactcatt tacctttcag caattaatga 1320
 aatatgctga caagggggca attagtagaa tttggcagct tgatgagtaa ttaaaattct 1380
 cttttgactt tgagccaggg tgtgtgacaa cagtctgtac aaactggtgt ccataccagc 1440
 aggtgggaag agctgtgtct ataaaaagcc aatgtccaag gtcacagagt tattagaact 1500
 acgtggaatc aatttttcac tgaagtagtc ctttttaca aaaagcaaac aaacatgggt 1560
 ctgttgtag gtaaaatgag ccgggttga tttatatggc attataaagc ttgtttacac 1620
 cttgcagtct gtcacctgct ttgaaggcac agccccgggc aacggggaga ggaaactgtg 1680
 actgacattc attgctactc catgaaatta tcaatgcctc ggtatttcta gcacttctcc 1740
 ctttatgaca aattaatgca aagtaatttc attagggaac tcgaggtaaa taatttgggg 1800
 ggaccctaag aggaagcacc tgctattaag gcaataggtg gaaagaagt taaagagatt 1860
 agaaaaaaga tcagtcacac accgaaagtc tggaggcttt gaatgtttc aaaattattt 1920
 ttcctatttc ctgaaattgc cctgcaattt cttaggcatt caggtagatg tcaggttagt 1980
 agctctcaaa tccttcacct cttcccatg atttcatgac ccctcccgcc accctgccat 2040
 tcatctagaa gaggtttggg tttatgctgc cccctcaga ctgaaaacac ctccagtcac 2100
 acagctctca agggaggcat ttctagtaat tgctttataa aatcctttca aatgtacaca 2160
 ttctcatggc acaacaatt acggaacttc aaattagcac tgctatattt atggatttca 2220
 atttatcacc cagaccagaa actgctgctg ctgctctctc tttgtaattt aaaacacgct 2280
 catcattctt ccctcttggc cggctctggg aagctgggtt tgcagcatct tgatcagctc 2340
 ttcggcagag ctgctgaaag gcagtgggag gagactttat catcagtga ccaaagccag 2400

gcctttcttc cgcctttggg attgggcaca agctgcctgt taaccatgta ccggtattca 2460
 aggcttcaaa acaaactcac acaattctgg gaaaagaaaa acatttctaa tctatttttc 2520
 aagtgataaa aacggcattt ctagtactta actgtacctg tcctgttttt taaatgggtc 2580
 tcagttttta accacatagg tattattttt tctataaaag ggggaaacta gaaaaactga 2640
 caactaaaaa aatagtaatc caagatatgc ttattgaata gctaatatct gacagaatac 2700
 tggacaaaat gagactac 2718

<210> 12
 <211> 355
 <212> DNA
 <213> Homo sapien

<400> 12
 gcaggtacac agttagtggg agcacactat ataaatcctt taacattgac accattcaac 60
 aatattttttt aaaatctaca aaatttttaa gtttcacttc ccatagcaaa atatcttcag 120
 tcaagaaatt agtctttgaa aattatgaaa atcggtgtgg gaaatattta tacaattat 180
 tacgtgataa tgcgacatat agtgtgaaac attgtgtcga gaatgcaatg agaataaac 240
 ctatttagga gataacccaa atgatttgta aaaaaattaa cttgtagaga agggaaggat 300
 gttgtgtaaa atcaagtcaa ttatttgagg tttttataat attgaatact tatgt 355

<210> 13
 <211> 969
 <212> DNA
 <213> Homo sapien

<400> 13
 gaccgaccaa tttttttttt tttttttttt tttttcactc taaagatact ttttatttaa 60
 atattttatg atgatacata tacaatatata atcttccaaa aaacaatgt aaaactaata 120
 caaatcactt tttcaggaac aaagaaaatc atttagaaaa tgtgattatg ctaaaagagg 180
 caggtaggt ttccaaggct gctcaagggt gaagcttaag accaactttt gtttgagtac 240
 acaagtgata ttacatttt catatactag tgatatgcct gttgcatact tggcaaaata 300
 aaactgatag taagtctata ataataaaag aaacaacaat tactaagtaa acaattctag 360
 atgatggaag agtaacctcc atttaagcta cagacttaga tgtctaaaaa tatgtgtcct 420
 gatctgtaca cagttagtgg gagcacacta tataaatcct ttgcatgaca ccattcaaca 480
 atatttttta aaatctacaa aatttttaaag tttcacttcc ctagcaaaat atcttcagtc 540
 aagaaattag tctttgaaaa ttatgaaaat tgttgtggga aatatttata caaattatta 600
 ctgataatgc acatatattt tgaaacattg tttctagaag caataaaata taacctattt 660

aggagataac ccaaatgatt tgtaaaaaaa ttaacttgta gaaaagggaa ggatggtgtg 720
 taaaatcaag tcaattatit gaggttttta taatattgag tacttatgta ctaagtcaca 780
 cccagccagt caataactga gaaattaaaa taaaataata atttcaaaga attacataaa 840
 tacagggcct tttagatit ttggcaattg taaacaaaaa cgaatggata gaaaaatact 900
 gtaagtatac gaaagatcaa tttagacca ggtagagcag aggtaacaca caagacaagg 960
 gcaatacgc 969

<210> 14
 <211> 470
 <212> DNA
 <213> Homo sapien

<400> 14
 gcagggtgctg ggcttgctg tggagggagt gacttgact ggcagcactg catgtcacct 60
 gggaaccct gcagacaaag ctaacatccc agacagacag atgtgaccag gacaaacgtg 120
 caataatgcc aaatgttaaa atgtgagttt accagcctag ctatgggact gctggctcct 180
 agtccaggaa tcatgggggt atgactgcct ctccaaccct gtgggctgta agcaagctca 240
 ggctagtctc cccactgggg gctgtgcccc tccctgggac ggttccgtgg gcagcccat 300
 cactgtgttc aatagtgtga gaatgtagct aaagccctg ctgctgctgc tgcacatgcc 360
 acagcaggcg gtgggggctg cgtggggaca atccatcgtg gagtgttctc tcagcttagg 420
 tctggacagg agacttggcg ggggatgcc caggatgtgg gtgattctgt 470

<210> 15
 <211> 1397
 <212> DNA
 <213> Homo sapien

<400> 15
 ggtgctgcac ctgtaccgga gcgggcagta tctgcagaac tccacggcaa gcagcagtac 60
 cgagtaccag tgtatcccag acagcaccat ccccaggaa gactaccgct gctggccatc 120
 ctaccaccac gggagctgcc tctttcagt gtccaacctg gctgaggctg tggatgtctg 180
 tgagagccat gccagtgtc gggcctttgt ggtcaccaac cagaccacct ggacagggtga 240
 gccagtggga gaagccctc caagggagat ggcaggacct ctctggaggt tgatagatag 300
 tgatccccca tcggaagtca gagggggtgc tgaggtgatg agagagaggt atacgtgtct 360
 tcaaggcagt caaattaggg agaattgtct tgcctccaga aagagaaaca tccagccctg 420
 ttacctctca cctctgcccc ccaggctggc agctggtctt tttcaagact ggatggagcc 480

11

aagtgggtccc tgatcccaac aagaccacat atgtgaaggc ctctgggtga cctatctgag 540
ggctcggctg accagctgac taccctcagc agctgggctt gcctgtggag ggagtgactt 600
gcactggcag cactgcatgt cacctgggaa cccctgcaga caaagctaac atcccagaca 660
gacagatgtg accaggacaa acgtgcaata atgccaaatg ttaaaatgtg agtttaccag 720
cctagctatg ggactgctgg ctccctagtc aggaatcatg ggggtatgac tgcctctcca 780
accctgtggg ctgtaagcaa gctcaggcta gtctccccac tgggggctgt gccctccct 840
gggacgggtc cgtgggcagc cccatcactg tgttcaatag tgtgagaatg tagctaaagc 900
ccctgctgct gctgctgcac atgccacagc aggcgggtgg ggctgctgg ggacaatcca 960
tcgtggagtg ttctctcagc ttaggtctgg acaggagact tggcggggga tgctccagga 1020
tgtgggtgat tctgtacctg gggaggctat ctctgacctc ccgacagggg aactcccag 1080
gccagcccag gggtcagggg cagaggtgca cacctcagca tgagccaaga ctggggtcag 1140
ggagcagggtg tggtttgagc caggacctgg ggcgggggtg gggccggggc ctttctgcct 1200
catttgcttt caatgaaagc ctcaaagcag ccaaaaccag gctttcccc ttctctgagt 1260
ttgaatatcc agaacttttt gtacttcttg ttggttaa at tgtttatttt tgtaaaaaat 1320
aaaataaaat tagttaataa aatgatgttt cacagcaaac tcttcctaa aaaaaaaaaa 1380
aaaaaaaaa ggcggctc 1397

<210> 16
<211> 680
<212> DNA
<213> Homo sapien

<400> 16
accaaaaaagc tgctgacagt ttgtgagcaa agttgtggat gacattatca gagctgtatt 60
ttaggaagtc ttaatatgtc aacatatgtc atactattat gttttctctc ccccgagtc 120
cattagccca ctgacctagg tgcctcttcc tcccgaaca caccagcatt cagcaattcc 180
ccaagggtccc tcccctgtct ccaaagctgt ctgcctgac actgacttag gcaaagcttc 240
ctacttttca gagacctgtg aaaggagacc aacccctgg ctacagccc ctagccctag 300
ttgttcccat ggacttctgt aaggatgtga ttcttttggc actcttcac tcctcccca 360
attcctgcaa gccctcagg agtggtgttc tcaatggtga cattgtgact ccaagccatg 420
aaatataggg cagttatcgc atcatagatg gattatatga gccttttatt ttcttcttgg 480
tgacaacggg gaacatggcg gcttcacaag agctgggaga gacagttgac tatacgtgtg 540
ctattactga agtaggtcc tcaaattgtt ggtggagcta ttggtgggtt gggggagggg 600

gttaaagggg aggccaggg gggaaggggg gccccggggg ggggggggaa aaaggagaaa 660
 agttttaatt tttccaaag 680

<210> 17
 <211> 1216
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (252)..(338)
 <223> a, c, g or t

<400> 17
 cccctaata aggcggtgcc cccctactgc ccttgaattt cgccttgaa tattgatgag 60
 tattggaatc tgcagagact ggataaaggt tgggatgagg tcgaacacta caggaacaga 120
 aaatatggaa catgtttggg agcaggccag ggattctgtc atataaagt catgaaaaag 180
 catatcatgt aatatttatg attattgctc tggagttaga ctgtttgggt ttgaatccca 240
 gatccagtgt tnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnntc ttggcttgtt accagaatta 360
 aatgagtttt tatgtgtgga gggctcatga gagtggctgt ccaaataag cattctctaa 420
 atgtagata tgactgtcat ccccttaaaa ctggcaggaa ggtagttga aaccatagca 480
 agccgagcca tgaatgccat gttaatgcat gttaatgcca ttattataaa ggtaccaaaa 540
 agctgctgac agtttgtgag caaagttgtg gatgacatta tcagagctgt attttaggaa 600
 gtcttaatat gtcaacatat gtcatactat tatgttttct ctccccgca gtccattagc 660
 ccactgacct aggtgcctct tcctcccgga acacaccagc attcagcaat tcccaaggt 720
 ccctccctg tctccaaagc tgtctgcctg atcactgact taggcaaagc ttctacttt 780
 tcagagacct gtgaaaggga gccaaccccc tggctcacag cccctagccc tagttgttcc 840
 catggacttg ctgaaggatg tgattctttt gccactcttc cactcctccc ccaattcctg 900
 caacccctc aggagtgggtg ttctcaatgg tgacattgtg actccaagcc atgaaatata 960
 ggccagttat tgcacatag atggattata tgagcctttt attttcttct tgggtgacaac 1020
 ggggaacatg cgccttcac aagagctggc agagacagtt gactatattg tatgctatta 1080
 actgaattat gcctcctcaa attgttgggtg gagctattgg tgggttgggg gaggggggta 1140
 aaggggaggc ccagggggga agggggggcc cggggggggg ggggaaaaag gagaaaagtt 1200
 ttaatttttt ccaaag 1216

<210> 18
 <211> 501
 <212> DNA
 <213> Homo sapien

<400> 18
 acagcattca tagaaatcac ataaggacac acttgaggga gtggaaagga gaaaatgatg 60
 gaagacaatt ttttttgaac tgaagataga aagatttctt tagattgaaa aggtcaccta 120
 agagccaaga agaaaaataa aatctagcat cattttgata acaattcaaa ctctaatagat 180
 aaaaaggaaa tcctagaagc tttcagagac aaaacaagat cattacaaag gaataagcat 240
 cggcattggg ggcacttcag caatactggt agtataacag ttctttccaa gcttgagata 300
 aaaatgattt caaatcttga atctgtaaac aaacaatcaa gtcattggtg gataggttgt 360
 caaatgaaa agattcagaa agtttacaac caatgggcac catatgaaga aaactgagaa 420
 tgtgcaaggt aaatacagaa gatactggaa acagtaagga ccaggatgac ccaggtgaca 480
 caattcttta gctgttactg t 501

<210> 19
 <211> 2418
 <212> DNA
 <213> Homo sapien

<400> 19
 tgtatatctg aaactactct aaaaaagtct cttaaaagaa agcaaggtaa ttttgttgtt 60
 gatactgaat gtaaggatca gtatcacaat attatttaat aattatgact gctagctaaa 120
 agaagatgga aaatgtttta aacactaacc cagaggtttc tgggttcagg aatagattaa 180
 gtaccataat ttgaaagaaa ttcattgggt cctgaggcag gtttctgggt tgggtggatc 240
 ctgagaaaga agtagaatag atcttgggt ccttcaaaat aatacagagg aaaattaaaa 300
 ggataggggtg ttgcactcat gggtaaaaaa ggctaaagca ctttgacttc agagtaaacc 360
 cctcttattt tgtcaaatgg tagcctgtc tgcctgttg tctgttccca ggccacctat 420
 cttacaggga actctgctg ttgacaagtg tcatgccttt ctatgaagcc taccctcttc 480
 ttcaaaagga ttgttaggga aacaggacaa ccaaactgca gatgcaactc acacaggagg 540
 aaaaagaata gaatggaaga gacagatcaa gacgaacaga cagaacaacc aacacctgga 600
 tgaaaaagaa acaatttagg taagagaaga gaatttaaaa aaaattaaaa ttctacttag 660
 tgtcttcggg agtattaagg aagttggatc cataaaacaa agatgactac taaacaaaaa 720
 gaagcaatta gaagatacaa aagagttctt ggaaatttaa atattcagta tacatgctaa 780
 atattagaaa gaacacagtt gaaaaaaga tcggcaatct gaaagataaa gtcacagaga 840

gataaataat aaaataaata gatgagaaac atgatggaaa agttaaggga catggtagat 900
 ttagtgggta cagcattcat agaaatcaca taaggacaca cttgaggagagg tggaaaggag 960
 aaaatgatgg aagacaattt ttttttgaac tgaagataga aagatttctt tagattgaaa 1020
 aggtccccta agagccaaga agaaaaataa aatctagcat ctttttgata acaattcaaa 1080
 ctctaattgat aaaaaggaaa tcctagaagc tttcagagac aaaacaagat cattacaaag 1140
 gaataagcat cggcattggg agcacttcag caatactggg agtataacag ttctttccaa 1200
 gcttgagata aaaatgattt caaatcttga atctgtaaac aaacaatcaa gtcattggtag 1260
 gataggttgt caaaatgaaa agattcagaa agtttacaac caatgggcac catatgaaga 1320
 aaactgagaa tgtgcaaggg aaatacagaa gatactggaa acagtaagga ccaggatgac 1380
 ccaggtgaca caattcttta gctgttactg tacttgattt tacaagaaaa atacttacgt 1440
 gttcattaca attacattgt aagggtgtt tgttttttagt tttaaaaatg aacctaggc 1500
 ccagagaata taattttagg tactgatcat aatgtaaact attataaact taacaatgta 1560
 agaaatcata acaataaaat ttcaatttaa aattcttgta gtttgtcatt aaattcactg 1620
 gttctgtgtg actttctcct atggcctatg ttggatatat aagagccaag tgttatccat 1680
 tcagattgct tcaatcatca tctcttcag ataaatgctg agtcagtatt ttcagctcaa 1740
 acgtcactct tagactccag cttatatatt agacagccag ctgaatgtct tgctctggat 1800
 gtcctaccag aatagcaaaa ttcaacatgt ctcagtctga atccatcatg tactccctct 1860
 taaatgctca acagcctttt cttatgatg cccctatctc taattaaccc taccacaact 1920
 taatacttag tctttgtatt agtcagggtt ctctagaggg acagaactaa taggatagat 1980
 gtatatataa aagggagttt attaaggagt actgactcac gatcacaagc tgagggtcca 2040
 taataggccg tctgcaagct gaggggaaag gaaggcagtc tgagtcccaa agctgaagaa 2100
 cttggagtcc aatgttcgag ggcaggaagc atccagtacg ggagaaggat gtaggccaga 2160
 agactaaacc agtctagctt ttccaagttc ttctgcctgc ttttattctg gttgtgctgg 2220
 cagctgatta gattgtgccc atccagattg aggggtgggc tgcctttccc agtcactga 2280
 cccaaatgtt aatctccttt ggcaacaccg tcatggacac acccaggagc gatactttgc 2340
 atctttcaat ccaatcaagt tgacactcaa tattaacat cacaacaact taaatgttct 2400
 caagtattaa aaaaaaaa 2418

<210> 20
 <211> 531
 <212> DNA

15

<213> Homo sapien

<400> 20

tacagagtat gtagtgggca tctgttgaat gaatgctttt cccagtagca gtgtattcat	60
acaatattaa tataattgtc ccctggctta cagataaaaa tgaaagcatc aagtgccag	120
tgagtgagac ccaggtgttc ttcctccacc cctagtggtc ccctgggcag gtcttttttt	180
tgtaacactc accagtctgt tctgtagtca atcattgatt gacttgctcg tgaacttgca	240
ggaactgttt catagtttca ttagcacaga gtaaacaagt ttgccatgca aggttatattt	300
gcatctgcat ttaagtata atgttgaatc aatgaaaagt gttgattaag cagtagttgt	360
agatatgcta agtttttcaa attactaata tcaagtggag attgttttta cttttaagg	420
tattgctttt gtgatagcat aaataatggt tttccttttt tgtaatgtaa attaattgct	480
ggcaactttt gtattcccat agactgggga agcttaattg cctttacaag t	531

<210> 21

<211> 1643

<212> DNA

<213> Homo sapien

<400> 21

ggcctttgca cattgaagtc ggcaactgctt tggcgctttt tttgtttttt ggctcgggtgt	60
tttgactgca agtctttttg gatagaattt tatagttaga aagtagctaa cacttgggtt	120
ttataggcac aaaaaacaag tcttatacta gctgtacttt attttttgag ttcttattaa	180
tgaggaacat ccacttttgc attgacagtg atttcaagat tgctttatca gcctttaaag	240
gattcttgac tagtcgtgca catcagaact gccaggctcc cagtgggtct gaagcagtaa	300
gctttgggtg ggctctggca tcagcacttt cactaagctt cacagataat tctgatgcat	360
actccaggcc tgaaccactg atcaatttga aacatgcata acaaagcaaa tcattcagag	420
agacaggctg ttgctccgga gtgatacaga tctggcagta cccagccctt gtgtgtgtgc	480
gttagctcag cacctgcca cactgcgagc ccccgtagga tgtgccttgt ccttccctgt	540
ttcagcactt aacacactac ctggtacaga gtatgtagtg ggcatctgtt gaatgaatgc	600
ttttccagc agcagtgtat tcatacaata ttaatataat tgtcccttg cttacagata	660
aaaaatgaaag catcaagtgc ccagtgagtg agaccaggt gttcttctc caccctagt	720
ggtcccttg gcaggctttt ttttttttgt aacactcacc agtctgttct gtagtcaatc	780
attgattgac ttgtctgtga acttgcagga actgtttcat agtttcatta gcacagagta	840
aacatgtttg ccatgcaagg ttattttgca tctgcattta agtgataatg ttgaatcaat	900
gaaaagtgtt gattaagcag tagttgtaga tatgctaagt ttttcaaatt actaatatca	960

```

agtggagatt gtttttactt ttaagggtat tgcttttgat atagcataaa taatgggttt 1020
ccttttttgt aatgtaaatt aattgctggc aacttttgta ttcccataga ctggggaagc 1080
ttaattgcct ttacaagtac ttatgtacaa ctttgatca aattttctgt aatagtttat 1140
gcttttagtac tataatgtga ctaataattt tatctgactt ctgtttatat catttgatca 1200
attacatggg tgtaaaactt ttctcaata tccttctatt tttatatatc tttctttctt 1260
tctattcctt tctaactctt attatattat tttaatctct ttcatttttt tctactctct 1320
tctcttctat ctttctaatt cagatttct actctattat attttttcta ttactccata 1380
tttatgtcta ttatcttatt ctaattatac tttttctct tttacttttc ttattatctc 1440
tccttctaac tttatctctc tttctttatt tgatcttttc ttttattttc tatattatc 1500
tttttttttt ttactcttct cttttatttg tcttatttct ctcaattatt catatttatt 1560
ctctctctta ctttctacat attcttactc ttatttttta tacctctctc ttatttacct 1620
tcctatcctt tcttggttct cct 1643

```

```

<210> 22
<211> 293
<212> DNA
<213> Homo sapien

```

```

<400> 22
acaaacatac cttgtttaaa ccaaccctta tcctgttaat cacctcttca cccaattaac 60
tacactagtt ccagctcctt tgtgttgatc tatttcacaa ttactactc tgtgtctact 120
tcagaacata agtgattatg tcatggagtc ttcttcctt aaagaatctc tcatgccaca 180
taatacatgt attaaataaa tttgtatgca ttttctgtt gatctgtctt atatcaattt 240
aattctcagg cttagcagag gatgaagaga actaggaaga tggatcatca aat 293

```

```

<210> 23
<211> 625
<212> DNA
<213> Homo sapien

```

```

<400> 23
ttttgcccc cccctctgcc ccccttttat gaagaccaga ttatgcaca gatttagccc 60
aagctgtttc tgctaggaga cctgcttctt cctaagaagc gtgctataga actggccagt 120
ccactctcca ttctcctagc cttggtattt tctggctgag agctttggat atgtcagcta 180
acctattcag cttattattt catttctaag agaggcataa caaggaaagg gctgtctctc 240
ctatttcaag ggattggggc aaacactaca ttagatttct gtgaatactc cttgtaaaag 300

```

17

cgtgagggcat aatacaaata tcagatatca gcgtgagttt tctatttcat tagacctatt 360
 tcattagaaa aggtgaaagc tctattatca ctctcttaat tgtttttagct cctttttgct 420
 tcaccttccc ttttatttct agtgtctact tggggcaatt aggcctcacg gctcatgtgt 480
 gtttgtgaaa aagaatTTTT aaatgtcttc tatttgctaa ggggaccatc cctactctt 540
 ggtctaagcg taatttctaa tcatataacc tgaagcatat tctccgatct cataaagtgg 600
 cattcttctg attctgatta gatgt 625

<210> 24
 <211> 739
 <212> DNA
 <213> Homo sapien

<400> 24
 ttttcgcccc cccctctgcc ccccttttat gaagaccaga ttatcgaca gatttagccc 60
 aagctgtttc tgctaggaga cctgcttctt cctaagaagc gtgctataga actggccagt 120
 ccactctcca ttctcctagc cttggtatTT tctggctgcg agctttggat atgtcagcta 180
 acctattcag cttattatTT catttctaata agaggcataa caaggaaagg gctgtctctc 240
 ctatttcaag ggattgcggc aaacactaca ttagatttct gtgaatactc cttgtaaaag 300
 cgtgagggcat aatacaaata tcagatatca gcgtgagttt tctatttcat tagacctatt 360
 tcattagaaa aggtgaaagc tctattatca ctctcttaat tgtttttagct cctttttgct 420
 tcaccttccc ttttatttct agtgtctact ttgtgcaatt aggcctcacg gctcatgtgt 480
 gtttgtgaaa aagaatTTTT aaatgtcttc tatttgctat gagaacatac cctactcttt 540
 gtctaagcgt aatttctaata catataacct gaagcatatt ctccgatctc ataaagtggc 600
 attcttctga ttctgattag atgtacagcc ctaatatcat agtgcaagta tacatgccct 660
 ccataagta ttctgaagta tgattcacc taggttttca aatctcttcc ttgccctaga 720
 aaacaaactt ggactcatg 739

<210> 25
 <211> 438
 <212> DNA
 <213> Homo sapien

<400> 25
 acaatatttt taaggacaaa aataacaatt atatacagtt gcaaagatca aattctaacc 60
 atggacacct ttcatctagt ccaatgactg aagcctgtcc aacgccagta actcccagg 120
 actaaggcca aatgaagcct caatgctgta agtttaccgt ttttgctgt tcacgatgct 180
 ttgttcttaa agaaacattt acgatttacc tgctttgaaa ctgtcaatag ctatattaat 240

aatgttttgt gccacaaatc aaagtccttt cctactcaaa agctactgtt aattgaaggc 300
aatgttacca ttgagatcaa attcagatgt ctagatccca gatacctggg tatgaaatat 360
gcaaatctgc caagagaaat tagatatattt tcttcttttc ttttaatatata acccactata 420
taatagtgaa ctaaatat 438

<210> 26
<211> 1706
<212> DNA
<213> Homo sapien

<400> 26
gtataaaaag gaacattgtg acaagaggca tatagccaaa ttaataggaa atttaagagg 60
aataaaagat tcccatttag cttgggatta accaaggcctt tttgaggaag ggagcattca 120
aagtgagtct ctgaagctga atcagacatt caggagactg ggtgaaaagt gtattctgag 180
gcgtatctgg attttctttt ttttttttcc tccctcttgc ctttgacaag gatcgcaaaa 240
gtggccgcac agccctgcat ttggcagctg aagaagcaaa tctggaactc attcgctctt 300
ttttggagcg gccagttgc ctgtcttttg tgaatgcaaa ggcttacaat ggcaacactg 360
ccctccatgt tgetgccagc ctgcagtatc ggttgacaca attagatgct gtccgcctgt 420
tgatgaggaa gggagcagac ccaagtactc ggaacttgga gaacgaacag ccagtgcatt 480
tggttcccga tggccctgtg ggagaacaga tccgacgtat cctgaaggga aagtccattc 540
agcagagagc tccaccgtat tagctccatt agcttgagc ctggctagca aactcactg 600
tcagttaggc agtctgatg tatctgtaca tagaccattt gccttatatt ggcaaatgta 660
agttgtttct atgaaacaaa catatttagt tcaactattat atagtgggtt atattaaaag 720
aaaagaagaa aaatatctaa tttctcttgg cagatttgca tatttcatac ccaggtatct 780
gggatctaga catctgaatt tgatctcaat ggtaacattg ccttcaatta acagttagctt 840
ttgagtagga aaggactttg atttgtggca caaaacatta ttaatatagc tattgacagt 900
ttcaaagcag gtaaattgta aatgtttctt taagaaaaag catgtgaaag gaaaaaggta 960
aatacagcat tgaggcttca tttggcctta gtccctggga gttactggcg ttggacaggc 1020
ttcagtcatt ggactagatg aaagggtgtcc atggttagaa tttgatcttt gcaaaactgta 1080
tataattgtt atttttgtcc ttaaaaatat tgtacatact tggttgttaa catggtcata 1140
tttgaaatgt ataagtcctt aaaatagaaa agaacaagtg aattgttgct atttaaaaaa 1200
attttacaat tcttactaag gagtttttat tgtgtaatca ctaagtcttt gtagataaag 1260
cagatgggga gttacggagt tgttccttta ctggctgaaa gatataattcg aattgtaaag 1320

atgctttttc tcatgcattg aaattataca ttatttgtag ggaattgcat gctttttttt 1380
 tttttttctc cgagacaggg tcttgctctg gcgcccaggc tggagtacag tggcatgac 1440
 ttggctcact tcagccttga cttgggctca agtgatcctc ctacctgagc cttctgagta 1500
 actggaacta caggtgtgca ctctctgcct ggctaatttt ttattttttg tacaggcagg 1560
 gatcttgca ctttgaccag ctggttttga cctcctgagc ttatgccatt ttgtgcctt 1620
 agtctcccaa aatgcgggga ttcccggagt gagccaccat gcccggttg cagttgcgtg 1680
 gaggagaacc ctctttatgg cttacc 1706

<210> 27
 <211> 387
 <212> DNA
 <213> Homo sapien

<400> 27
 catttgccaa cataccattt ttaatggaga ctcaaaacat taaaaaaaaa aatcagaact 60
 gagcattgcc aggagaggtc agacttgcca taggatagac tttctgggtc tcatatgaag 120
 cctctacaga cagaagcgtg tcctatgttc atggcctttc tggatgtaaa ctggagtctc 180
 tgacaaacta cagtgetttt ccaagctcac ctctctagcc tgtgatgaac actgtcgaat 240
 acattaagtg aaacaccaa gcttagaggg tgctgagcaa cagaaaatgg gtatcagttg 300
 gtccgcattc ggacctogta ttcgtattga tggttctccc cctccttgcc tcctccctac 360
 tccacctctg ctgcccttat gcttggt 387

<210> 28
 <211> 873
 <212> DNA
 <213> Homo sapien

<400> 28
 cagggacgag tccccagaac cacagcgccc aaagttgggc caggtccagg cactgcgaat 60
 aatgtgtgaa gagtcatcca agttagactt ctctgaattt ggagccaaga ggaagttcac 120
 cagagcttta tgaggctctga agaagaggga gagaaagaga ggacagaaaa cagagaagaa 180
 gggagggtttg catctggacg gcggtcccag tatcggagaa gcactgacag ggaggaagag 240
 gaagaaatgg acgatgaagc catcattgct gcttggagac gccggcaaga agaaaccagg 300
 accaagctgc agaaaaggag ggaggactga gctggggaaa atctgagaac actgaaagaa 360
 accactcacg ttagcatagg gctcagggca cacgttgcca cactcatcg caggatgagg 420
 atacagagag gatcttccag aggggcagag ccaaaatgag aggtaccaag cataagggca 480

20

gcagaggtgg agtagggagg aggcaaggag ggggagaacc atcaatacga atacgaggtc 540
 cgaatgcgga ccaactgata ccattttctg ttgctcagca coctctaagc tttggtgttt 600
 cacttaatgt attcgacagt gttcatcaca ggctagagag gtgagcttgg aaaagcactg 660
 tagtttgtca gagactccag ttacatcca gaaaggccat gaacatagga cagcgttctg 720
 tctgtagagg cttcatatga gaccagaaa gtctatccta tggcaagtct gacctctcct 780
 ggcaatgctc agttctgatt tttttttttt aatgttttga gtctccatta aaaatggat 840
 gttggcaaaa aaaaaaaaaa aaaaattgcg gtc 873

<210> 29
 <211> 159
 <212> DNA
 <213> Homo sapien

<400> 29
 actagaggat gaaaactgaa acgttggttt gatgtttatt gaataacgag attagagaat 60
 atttgatttt tgtgtcagt gtattaaaga aattttcaca ttgataaatg ttctctagga 120
 atgtgtctac attcatcagg tgtgaactct tgtacctgc 159

<210> 30
 <211> 1832
 <212> DNA
 <213> Homo sapien

<400> 30
 ggcaggagaa ctgcttgtaa cctggggggc ggagggttgca gtgagccggg atcgtgccat 60
 tgcactccag tctgggtgac agagcaagac tcattctcaa aaaaaaaaaa aaaaggaatt 120
 tttattacta tttcctgaag aatgggtttt gttaacttgt tactgtatca ttaaaaagac 180
 cttctaattg ttcagtacaa taatctagaa cttgatttat gtggcttttt atagttatct 240
 gaatgcattc cttttgccac atagaccata tggctagttc tccaactttt ttgcttattt 300
 ttaataaacc ttgctgttca acaatcagag aaacctttag attttgatg attcttccag 360
 ttgaggtaga aacatcttag ataataggaa aggcaaatc aaagtcctaa cattttcata 420
 gtgaggttta caagtaaaat aacttatcca tataggttat cttcgttggt tagcaccagt 480
 ataaatagtg atttcattaa tcattgaatc agatgaagca gttataaatc actttttact 540
 ttgtgctaag aattattgta atttcaggac actttattat ttctctgag cagtttccat 600
 tggaagggtg agtttccctt ttttaagttc taatcatcac taaagggtta gataatcaaa 660
 taggaggtta aataagttat gtttgatctt tttcccttga aaataatgct gaacttattg 720
 tctacattct gattattagg cagaaatgca cttgtttaaa tcatagaagt aattcatttg 780

gaggatataa ttactcgatt ttctagtggg gtgaaatact ttttaacaat tgtgcttgct 840
 tgtaactgaa atgttataaa attttaacac tatagggatt atagagttat attagctctc 900
 ctcaagagac tgaagcacia tatttttcat gtaacaattc ttatccaagt gctgctaate 960
 tgtcgtgcaa ataataagc tatttggttg cctatttagc tattcacaaa tcaactgtaat 1020
 ctttgaaaca atcttgctgt tcatttgtat taatatttgg atattgtgag ttaatacttt 1080
 agaaaaaaat ccatcaactc agccccgtta gcaaaactgt ttggattcat agtttttata 1140
 tgtgttaaca gtagaataaa ttttgaaggg gctatttact accaatgact aaggggaaaa 1200
 ttatactgtc actatcattt gacttgaaca tttgtggtat tgtaaaagtc ttgtcagttg 1260
 tgttctaaat tgcttaagcc atacgttctc ttaaacagga tgtttttttc ttcctttcca 1320
 gcagcctttt tcttctttgt ctgttatggg taatactcca tagattttag aaattgagaa 1380
 gttcttgaaa cattttattt tcttgagttc atcacttttg actcttgat gagatgtgat 1440
 ttgtcataaa agatagcctt ccactacttc actaaatgaa tttcagagta aacactgtga 1500
 ttctgcagag cggattcagt aggctttcca atgttttctc ctgctataca gtgcctacca 1560
 ccttgagggc acttcagtag tagaggatga aaactgaaac gttgttttga tgtttattga 1620
 ataacgagat tagagaatat ttgatttttg ttgtcagtgat attaaagaaa ttttcacatt 1680
 gataaatggt ctctaggaat gtgtctacat tcatcagggtg tgaactcttg tacatgaatt 1740
 ttgtaccttg aatccacata tatattaagt gtatcatcaa tataaaaaata aacattattt 1800
 gcttaaaaaa aaaaaaaaaa aaaatggcgg tc 1832

<210> 31
 <211> 531
 <212> DNA
 <213> Homo sapien

<400> 31
 actcttagta tactatgtgc ccttgtagtc ttttctttcc tccatattca agaaatccat 60
 gatagagtat taaaataatg ttctaataaa ctccctgaat tcattcacat gtattgtatt 120
 cacttttata ccacatctgc ttttacagtt acaaacattg aaaatatact accctcaatc 180
 gagcttcaca tgctgttgct atcagtttgc taagacttaa agaataaaat aataggctaa 240
 ttcttttaaa catcaaatgt gctcttaggg ttaatttgta atctttaatt catctttcac 300
 taaattttta agatattttt ttgctcccc tatagatctc atttctatt tcaatctgaa 360
 atgattttct ttaaaactgg ttatccgtta tggaatatct ctgcataatt aaccatttct 420
 ttctctccct ctcttataaa ataataattt gttttatgaa tcattccctt ttatttttaa 480

tcttcaattg ctctttctcc aacagatcct tcattccact ctctaatagt t 531

<210> 32
<211> 1001
<212> DNA
<213> Homo sapien

<400> 32
ggccggcggt aaatccttag ggtaatcctg tcccttaaat atctccggta ctcttagtat 60
actatgtgcc ctgtgtatgc ttttctttcc tccatattca agaaatccat gatagagtat 120
tāaaataatg ttctaataaa ctccctgaat tcattcacat gtattgtatt cacttttata 180
ccacatctgc ttttacagtt acaaacattg aaaatatcct accctcaatc gagcttcaca 240
tgctgttget atcagtttgc taagacttaa agaataaaat aataggctaa ttctttāaaa 300
catcaaatgt gctcttaggg ttaatttgta atctttaatt catctttcac taaattttta 360
agatattttct ttgctccccc tatagatctc atttcttatt tcaatctgaa atgattttct 420
tāaaactggg ttatccgtta tggaatatct ctgcataatt aacccatttc ttctccctt 480
ctcttataaa ataataattt gttttatgaa tcattccctt ttattttāaa tcttcaattg 540
ctctttctcc aacagatcct tcattccact ctctaatagt ttgggttaatt ctttatagta 600
actgctctcc cagcactgtg gcagacactg gacctactat acgāāāacta tcaactaacc 660
cttcttctt accttctcc acaataaaga ctagcāagcc aataactcaa ctgtacattc 720
tcccttgag tcagaaatag ccattctaca tgggtgtgac cactgtaaca ttgctagāā 780
ccctgcgga gagattctgt cattāāāā acaggagagc ttgccaggag āāāacttg 840
tctccaccac ttccacattt tctgcctgga atgtgggtta gcctgggtgga gcagcactgt 900
cttgcaacag taagttgtta ccttaagaga aaggtgtaat gctacāāāag gtatgāāagc 960
attagagact ttgatataca gāāāagatat tagāāāāagc a 1001

<210> 33
<211> 420
<212> DNA
<213> Homo sapien

<400> 33
actttttgca tttctacatt cagataāāāā gatttgcattg cacctggcta acgccaaggg 60
aacttcattt ttttcttcac tattatgcac tttcatggta tagtctttct cagttctttt 120
aattttgtt atttaacatc tttaatagca cagcāāacat cttttcagāā attttcagtt 180
aaagcctttg aattacttat ctttgattta atttacagcc agcattttgc cagttctāā 240

23

ataatatatta gctcaactga ttcatacgta ttaatgacca ttctagcaaa ggcctacaag 300
 tgggtgtggga atcagggaaa ggctgcctct ttggtatctc aactggtatt gattattgct 360
 atcaactatt tggggagaaa aaatcaaaat gaagccctgt caaatcttag aagtacctgc 420

<210> 34
 <211> 1613
 <212> DNA
 <213> Homo sapien

<400> 34
 cgtacatgac atgaataaat tcccatgctg ttttgggtatt agtaataaca gtgactacgt 60
 ccgtgtctta gtatagcgcc ctgcgagat aattacggcg tagttacttg gagaatatgc 120
 acccgtttgg ggattogaac atacatgggt aaagttaatg tgggaaactc acgttaagat 180
 catgggagac attgggtttc agaacatgta atatcccggt tgcacccagt ttaacagccg 240
 tcttaattgg cctgaaagcc aaaaatagac tttctgaaat accagattag ttaaaaatac 300
 tttccattga tagcagtgtc agtccctaga acaaaaggta agcaaaactt atttgaagt 360
 tactgcctat tcaatgcca gaatatgtag atcctaaatc taagccctta atatacatct 420
 actttaaga taactgaaag atctcacatg cctgataatc cttaatttaa accgtcctgt 480
 aaacatagtc aaaatctgct aatagaaata caattcaagt aaacattgca tatttgattt 540
 aaaccacctt acagttaaatt tcaatcatga cacattggat cataaccact aatatgtaaa 600
 aagttttaa aaaaatcatc cttaactgata gatgaaaata aactttgtaa acttggtcat 660
 ttaaaaataac gaatgtactg cagctgctct ttggtttggc atagtttcag gtactgaata 720
 ttcaagtaaa tttgttccca ggtaaacc aa gtctcctaatt ttgtctgtaa tggcaatggc 780
 aagacctgaa cttcaacttt atttttctta aggtgtcatc acaaagtgtt tgaaggacca 840
 aagatagtac ttctaaaatt tgacagggct tcattttgat tttttctccc caaatagttg 900
 atagcaataa tcaataccag ttgagatacc aaagaggcag cctttccctg attcccacac 960
 cacttgtagg cctttgctag aatggtcatt aatacgtatg aatcagttga gctaaatatt 1020
 atttagaacg tggcaaaatg ctggctgtaa attaaatcaa agataagtaa ttcaaaggct 1080
 ttaactgaaa atttctgaaa agatgtttgc tgtgctatta aagatgttaa ataacaaaaa 1140
 ttaaaagaac tgagaaagac tataccatga aagtgcataa tagtgaagaa aaaaatgaag 1200
 ttcccttggc gttagccagg tgcattgcaa tctttttatc tgaatgtaga aatgcaaaaa 1260
 gtaccaggag aacatttctg aaagtagtca agtatgtttt aacatttatt tccttataat 1320
 atgcaaactg ccaaactgga gttatgtttt tagttggtaa ttgatataata tatatatttt 1380

24

tgagatggag ttctactcgt cgcccaggct ggagtgcagt ggcacgatct cggctcactg 1440
 cgacctocac ctccctgggtt caagtgattc tcctgcctcc acctcccgag tagttgggac 1500
 cacaggcgtg tgccaccatg cctggacagt tttggggttt ttttgtattt ttagtggaga 1560
 taggggtttt ccactcttgac caggctaata tcgaacctc gtgccgaatt ctt 1613

<210> 35
 <211> 597
 <212> DNA
 <213> Homo sapien

<400> 35
 acctattcac cattccaacg tgaagaagct ctgcagtagg aaaaataatt aacacactta 60
 tagtctactg cccatgtaag gatcagctcc ggctaagagg ccaaagatgg gtgacatcgt 120
 tatgctctgc ctttattttt tctttcttac ccacttagct tcctaattgg aggaaggagg 180
 cgtggtaaag gtatatgaag actatggctt aattagacca gaaaacactg tcataatctc 240
 tggggtcac agaatgtcca gttttgtctt tgggccaaga taagggcagt gggattttatg 300
 atgtgttgtt tatagtctga aactactctg gtgatcacca gggtcagttt ctttaatgat 360
 ggtttccaac tggcctaata cattaagtaa gactggctga taacatgacc agacagacat 420
 aaagaccctg ttgggaatga cattgaactc tcaaagtcaa gatttcttac acaaacttat 480
 cagctggaga aatgaaggc agtgtggtat atgtgtgcca ataaggacat tatgaagctt 540
 aaatatggaa tgtctcttgg acccccgatg tcactctgtat tctctttttc ttcttgt 597

<210> 36
 <211> 1327
 <212> DNA
 <213> Homo sapien

<400> 36
 ggaagacctg attgggaata gtcgaaagcc ttgatatgtg caaagaaaga accatttgat 60
 caaccagtt cttaatacag gatactaact taaaatatag actcaagtta tacgataatt 120
 caaacattta ttgtatttat actattctat atgtactttt ccaggaacca ggaatacaaa 180
 actgacatgt tctctgtaca gaggtcaga ctagtagaga acagttaggt acgccgttaa 240
 ttataaacta atatgtatca tcaattatgg gtttttatgg gggtttggca ggtggaaggg 300
 accagggaga gatgatgagt gatgatggtt atgtagtctt taggaggatg caattataac 360
 attgctcttc ctttcacgca ccacatgatt tagcaagtac ttcatattgg ctccaccatt 420
 aacatggcca atggcttctg gatactcaca gttcaggcac agtttctcct gaagattttt 480
 tacctctccc atctttaaga aattgtctgg atgtccatga aagatgctga cacttgatt 540

aattcattaa aaaacaccac cccctccctg aaataaacta aaaagtaatg aattcataga 600
aaaaaatttc accaagattg aaactagaga atatacctag acttgcaactt tgagctttga 660
gaaatgtgta cctattcacc attccaacgt gaagaagctc tgcagtagga aaaataatta 720
acacacttat agtctactgc ccatgtaagg atcagctccg gctaagaggc caaagatggg 780
tgacatcggtt atgctctgcc tttatttttt ctttcttacc cacttagctt cctaattgga 840
ggaaggaggc gtggtaaagg tatatgaaga ctatggttta attagaccag aaaacactgt 900
cataatctct ggggtcatca gaatgtccag ttttgtcttt gggccaagat aagggcagtg 960
ggatttatga tgtgttgttt atagtctgaa actactctgg tgatcaccag ggtcagtttc 1020
tttaatgatg gtttccaact ggcctaatac attaagtaag actggctgat aacatgacca 1080
gacagacata aagaccctgt tgggaatgac attgaactct caaagtcaag atttcttaca 1140
caaatctatc agctggagaa aatgaaggca gtgtggtata tgtgtgcaaa taaggacatt 1200
atgaagctta aatatggaat gtctcttgga ccccgatgt catctgtatt ctctttttct 1260
tcttgtacta tatcctttgc ctgtáaataa aaggtttatt tgaaaaaaaa aaaaaaaaaa 1320
gatcggc 1327

<210> 37
<211> 172
<212> DNA
<213> Homo sapien

<400> 37
acagagcagg ggtcagcaga tggattttgt aaagcatcaa cttgtaaata ttttcaagtt 60
tattagctgt atggctctgg tttctgttcc ctgttccaaa tgttaaagtc tactgttgta 120
ttctaaaagc agccatggac tgaatgtagc tgtgttccaa taaaacttac ac 172

<210> 38
<211> 1547
<212> DNA
<213> Homo sapien

<400> 38
gagcaaaactg cccttcatct actgtggata tgttggggga tgatggaata tagtgaaaga 60
taatgggtgc tcatacagca gtctagactt aagggtattc aactactata tattaacta 120
gattatcttt tatttttttaa ttttgaaatc tggatgctca agctctgcct gcacaaccac 180
atgaggaaga aggaacaatg acaacaaaaa taactactaaa tttaaattta agagtactac 240
ttttaggaaa tagacaaacc attatttggg tacaactaaa ggcaactggc atggactcaa 300


```

atattttggg gaagaaaaag actaaaagtt ctaaggaaga aaatgcgaac cttgatagtt 360
tgaaatagtt aaaaagacag tgtagaaact gtttaggcag ttgattatg gactattaga 420
tgatacttgg gtctgataat ggtataagga gaataaagta tttagggatc caatattacg 480
cctgcagctt tttccaaata gttcatgggg gaggggggatg atggaatata gtgaaagata 540
atgggtgctc atacagcagt ctagacttaa ggtgattcaa ctactatata ttaactaga 600
ttatctttaa atttttaatt ttgaaatctg gatgctcaag ctctgcctgc acaaccacat 660
gaggaagaag gaacgatgac aacaaaaata acactaaatt taaatttaag agtactactt 720
ttagtaaaaca gacaaacat tatttgggta caactaaagg caactggcat ggactcaaat 780
attttgggga agaaaaagac taaaagttct aaggaagaaa atgcggacct tgatagtttg 840
aaatagttaa aaagacagt tagaaactgc tttaggcagt ttgattatgg actattagat 900
gatacttggg tctgataatg gtataaggag aataaagtat ttagggatcc aatattacgc 960
ctgcagcttt ttccaaatag ttcatggggg aggggggatgt gtaagtgggt aactgaagtc 1020
taactagata ggtttgttgt aagcttagga tgtttacagt tcttcatgtt aagttgagcg 1080
tgatgggaag ggaagaatg ctgatcttta aatttttgtc cttagttaag ttctgtattt 1140
agtgaattaa ttgcatccta aaaagtcaaa cttgaaaagc acatttttaa tggcaaatct 1200
attttttaca tgtttgtgaa gtttttattt ttagtaaac agaccatcag aagagaacaa 1260
tggtacagag caggggtcag cagatggatt ttgtaaagca tcaacttgta aatattttca 1320
agtttattag ctgtatggct ctggtttctg ttccctgttc caaatgttaa agtctactgt 1380
tgtattctaa aagcagccat ggactgaatg tagctgtgtt ccaataaaac ttacacaaaa 1440
gcaggcagtg ggccataatt tgcaacacct gattcacagc ataattttgt cacaaactga 1500
aagtgttcct caattaaagt gatttttttt tcttgaaaaa aaaaaaa 1547

```

```

<210> 39
<211> 360
<212> DNA
<213> Homo sapien

```

```

<400> 39
agcaaagtcc tcttctatgt ggttatctgg gactcctttt ggaggaaca ttttaaattt 60
tccatttcaa agcattctgt tggccttctt acactgtttt tctctgccta tcttgggacc 120
tgagttctcc tggacatgaa tctgcagcca cagagcctag aagctcatte ctccacatte 180
tgtgactgtt ccccaaacac agggagaatt tgcagaaaat aagcccaaaa atcttgccat 240
tctttgcaat aaaacccac attacaaact gctgaaaaca ggattttagc ctgaataggt 300

```

tgttcctcta tttgaaagcc tttacaattt cggagggaag tttccaaatc atcagtaagt 360

<210> 40
 <211> 754
 <212> DNA
 <213> Homo sapien

<400> 40
 gtgaaaacaa acccactgag acccgtctg ggttttotca gaccctaaaa tctgatcgaa 60
 taatgatagc gttcgtacac attcacctcg gcctgtctta agattcaaaa actttccaag 120
 actctagggg aatctttcca gacgctagac ccgagttaaa gattagatgt tgattgaatg 180
 aaacactcct gcttgtaggt gcaatccac atggagctta agatatatat aagcactaga 240
 aaaaaaaact tgtaactttg agttgatctg gtgatttacc tggcgcttct ccctgtaagt 300
 ggctgcagaa ataaacttcc ttctttccca gtctgtctgt atcttagtat tgaacaattg 360
 cgatggagct gccagcaaa gtccctcttct atgtgggttat ctgggactcc ttttgagggg 420
 aacatthtaaa attttccatt tcaaagcatt ctgttggcct tcttacctg ttttctctg 480
 cctatcctgg gacctgagtt ctctggaca tgaatctgca gccacagagc ctagaagctc 540
 attcctccac attctgtgac tgttcccaa acacagggag aatttgcaga aaataagccc 600
 aaaaatcttg ccattctttg caataaaacc ccacattaca aactgctgaa aacaggatth 660
 tagcctgaat aggttgthcc tctatttgaa agcctttaca atttcggagg gaagthtcca 720
 aatcaatcag taagtacccc ccactccagg ttha 754

<210> 41
 <211> 635
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (560)..(579)
 <223> a, c, g or t

<400> 41
 ccgcccgggc ggtacctatt tgtaatcadc agagtatata catctgatta ggactcagct 60
 atgttcaagg cttcatcgag cccacatac aattatcatt tgcattttct gctacaatcc 120
 aagaaaacac cttgtgtgct attagtggcc cttgcaagaa ggaagatgct gttttccata 180
 acaggaaatc aacgaacgaa caaagataat ccgtctctcc atcttacaaa aacaaagaaa 240
 gcctagcaga aaagtgaac aggacagggt cctgaaaaac atctagtgat gccataaca 300
 tggaatgtht tthaaaaagt gatttgtctc actgaagctg cagaagggtg tcccactt 360

atatattatg tgactgcact aaaaacagac gcttttggtg cactgagcgt tacaaaaagg 420
 cagaaagctc acaaatagat gcaatttttag gtatgggaat aaaatgacat aaagaaactg 480
 accttggtat cagtttattc tgtagagtgc aagataagga tattccaagg aaaaacctat 540
 tacaggtagt atatagagtn nnnnnnnnnn nnnnnnnnna agccgaatcc agcacactgg 600
 gggcgtaacta gtggatcgag tcgggacaag ttggg 635

<210> 42
 <211> 1142
 <212> DNA
 <213> Homo sapien

<400> 42
 tttttttttt ttaaagtttt acttggaata tgtgtatttg cttaaagtac aagggaat 60
 attgcaaatt atacatcatt tgaaaaatta tctctcttta gttaattttc agtcacaata 120
 ttggatgtag cagctccaaa tagaggttac ctgattattg cttttataat tgaattctta 180
 aagagtttac atcataatta tataattgta tttttgaaac atcacagaaa cccaacatgt 240
 acctatttgt aatcatcaga gtatatacat ctgattagga ctgagctatg ttcaaggctt 300
 catcgagccc aacatacaat tatcatttgc attttctgct acaatcaaag aaaacacatt 360
 gtgtgctatt agtggccatt gcaagaagga agatgctggt ttcaataaca ggaaatcaag 420
 aacaaacaaa ataatcgtct tccattttaa aaaaaagaa agcctacaga aaagtga 480
 ggacagggtc ctaaaaacat ctagtgatgc caataaaatg gaatgttttt taaaaagtga 540
 tttgtctcac tgaagctgca gaagggtatc ccacacttat atattatgtg actgcactaa 600
 aaacagacgc ttttggtgca ctgagcgtta caaaaaggca gaaagctcac aaatagatgc 660
 aatttttaggt atgggaataa aatgacataa agaaactgac cttgttatca gtttagctgt 720
 agagtgaag ataaggatat ttcaaggaaa aacctattac aggtagtata tagagtactt 780
 gggcccagtt gaagcccagg taatgtgatg atagtaatga taatggtcca ctgaatgcta 840
 acagacaagt atatatagtt acagctgtac atggatatca caaccttaca cacaattct 900
 agaaagatca ttgtgaaaat gacattccat aaatcacatg gaatcagcac caagtgtgtc 960
 tttatgcatg cccaaaaagg aaggagaaac tgacaacat caataatgaa caatgactta 1020
 tttcaaattc aatatctagt gctgataaat ttattttggt gttgttgttt aaacgagaac 1080
 gtttctatgg gcctcctaag tcattctatg ctaaaaaata acagctcttt tttgtgtct 1140
 tt 1142

<210> 43
 <211> 498
 <212> DNA
 <213> Homo sapien

<400> 43
 gccttactgt atcaagcttt tataatgatg actccttcat tatttaaatt cctatacttt 60
 tatttgttat cacgcaacta ctttgttcaa tgtgaaaatg tgctaactca tgggagaaga 120
 gtgccaattg atagttcttt tagcaattaa gaatatggtg tttgggaaga aaagtttgaa 180
 atgcaacaaa tggatatctt aacacagtag tattatatta tcagttcttt agtaagtgat 240
 tttagagatg ttgtaggcta cttttacggt ggaatatata gtatagagat gcaaaactta 300
 aatgtttaca tcaatttata ttgaatgtca cataatttca tggaaggaaa ggtagcttga 360
 tatttagatt ctaagatata atctgaaagg aaactaatta tgttctctac acttactgta 420
 atactgatta ttcttacata tcaaattatt gaactttaaa aatttcattg tatagtcatt 480
 aaactgagtt gggttttt 498

<210> 44
 <211> 2254
 <212> DNA
 <213> Homo sapien

<400> 44
 gagggtctgtg gcgcgatctc ggcttactgc aacctccac tccctgggtc aagggattct 60
 cctgcctcag cctctgagtg gctgggattg caggcgtgag cactgcgccc ggctatact 120
 gtatatattt ttaaagactg ttctaataga tataaaaact gtaaaaaata agtattttta 180
 tatagctctc atggatttta ttaaacagaa ttggctcaaa aatactatgt tacagactgt 240
 tgggtaccct tgccaacgt gaactggcag tgttaccttg cttttgcagt aatagtctac 300
 agattgcagg tctcatcaat tccatccaaa gtttaaaagc atttaaaatt accaaatctt 360
 taaaatcact ttggtggtga ttccaaattg gtaccaagca aactttcttg atgccaaca 420
 tgattttcag taaccaccct ttagagtatt tgtttactaa gttcaccaca ttttgaacat 480
 ggtagtttta gactgcaata atatttagac ttacattatt acttactgct aagtaaaatc 540
 taaatcctgc aaatgcacag aattcaagct gaaatataat gatttatggt tagctcacat 600
 tgaagtattg gttggttact tatgtattaa tgcagtgtgc attcacattt aatcaggttt 660
 agtctgtttc tattttaata attttaaaaa attatacaag caaattagat attagacatg 720
 ttagttacaa tggtaacaca tttttagggtg tcgaaacaca attttcaaaa ttctaataga 780
 aagttataaa aatgtaaaca agaattgtaa aaatggacaa agtagtcaaa tatattttca 840

30

```

aagcacaatt ttattagaca ggcataattt acatttttget tttctagtgg gtttgaaaat 900
gtttattgga gattgggcta tgtagtttat aatttttaat tcataaaaaa gtaatcatat 960
atgagaaggt agacctgtgc cctaggatca tgtcacatat acagataatg ccatttcctt 1020
gtgtgtgtga tgtgtgtttt gatgacctcc acaggcctta ctgtatcaag cttttataat 1080
gatgactcct tcattattta aattcctata ctttttattt gttatcacgc aactactttg 1140
ttcaatgtga aaatgtgcta actcatggga gaagagtgcc aattgatagt tcttttagca 1200
attaagaata tggatatttg gaagaaaagt ttgaaatgca acaaattgat atttcaacac 1260
agtagtatta tattatcagt tcttttagta gtgattttag agatgttgta ggctactttt 1320
acggtggaat atatagtata gagatgcaaa acttaaatgt ttacatcaat ttatattgaa 1380
tgtcacataa tttcatggaa ggaaaggtag cttgatattt agattctaag atataatctg 1440
aaaggaaact aattatgttc tctacactta ctgtaatact gattattcctt acatatcaaa 1500
ttattgaact ttaaaaaattt cattgtatag tcattaaact gagttgggtt ttttcttaaa 1560
gggttttagca tcaactcattt gatttacaca ttcacattat aatatttaaat tatcatgggt 1620
gtatgcttta cataaaaaag gtttataaaa gttatttatg ctatattgaa agtcatctta 1680
agaatctoca ggttatttaa agtagttata ggagcagaga acaagcacct ttatcaaaat 1740
ctggctctat gtgccttgct ttaccaaata cctgattttt ctggagggtg ttctctgaat 1800
tcacaactgt agacacatgg gcaaaattag gatttttaag aataaatada tttctatttt 1860
tttggttggt tcaacattag ctcttcaaat tcattaacaa aattaaaata ggtatattac 1920
aaaagcataa acatttgtga acagtactta aataaattgt gatactattg ctccatcatt 1980
gaactttttg aaactttaac aattgtataa aactgtcagt ttgttgtttc atttgtaatt 2040
acaaaataat ttaaaaaactt tttaaaataa tttggatcct gactttgtct atatctgtat 2100
ttcatttggt tagaaagatt cttttgggtt tgataatgta atttgatat ttaaattttt 2160
tatggacata attcaaagga atgtataaat tggctctttg ttaaattggct ttttaattga 2220
aaaaaaaaa aaaaaaaaaa aaaaaaatg gcgg 2254

```

```

<210> 45
<211> 573
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (310)..(498)
<223> a, c, g or t

```

<400> 45
 ttgcgcgccc cccggcagta ctacatatcc caccaccagg agggaaaagc cactgggttaa 60
 agaggaaaat ggggcaccca taccgctctt cgaacgggtt aaaaaatggt tatgaaggac 120
 attattgtaa taactgacaa aatctgaata tgcactgtat attcatattt gataatagca 180
 cattaatata agataccctg aatttggttaa ttatattgtt ggtaagagaa taatcttctt 240
 agggaaacata agctgaagta tctgaagtta aatggatatg gtatttccta tctactcttt 300
 tttttttttn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
 nnnnnnnnnn nnnnnnnnct ggtcaggggt gctaaaatat atgtcggatg ataaggcatt 540
 ttgccaatg tcacaaacat gattcgggta caa 573

<210> 46
 <211> 537
 <212> DNA
 <213> Homo sapien

<400> 46
 ccgcccggcc aggtacctta ataattgttc atcagggtcaa aatctatcct gtccctctagg 60
 aattctgggc ttccctcagg cctagcagag agctttctgc cactactcag gcaaccaagg 120
 gtgaagtgct tcaagtagta tttgtggaca gcagcagggt accattgtga ggtagatatt 180
 ttgttctaatt ttccagatg aggaagctga gaccctaaaa ggctgaccgg ttccctgatg 240
 tgttacctgc ttctgtact gatccaaact gcagaacttc tcattcatcc ccaaggcctc 300
 caggcagtat ccaatgggga atcagctcta aaaggaacca gaccaacgtt ttccagcccc 360
 ttcatctgt agcttcctc tgtgtgagga aaggatagaa atgttcagga catcatcata 420
 caggctcctc atctacaaag ttccagtagc agtgacgcct acacggaaga cttggaactg 480
 caaacaggct ggggtcacct cagtgcacac tgacgctgtc caaccagaag ttcgatt 537

<210> 47
 <211> 797
 <212> DNA
 <213> Homo sapien

<400> 47
 aaggtcagta aaacaaaaag ctagcagagg gcaggctcag gccctggggt agagggctaa 60
 ttaacttctg tcagctagtt gaatagagcc ttgtgtgctt gttagagacc aaaggctactt 120
 caaaggaaaa aaatctagat tcttcctgt gtaccttaat aattgttcat cagggtcaaaa 180

tctatcctgt cctctaggaa ttctgggtctt ccctcaggcc tagcagagag ctttctgcca 240
 ctactcaggc aaccaagggg gaagtgtctt aagtagtatt tgtggacagc agcaggtgac 300
 cattgtgagg tagatatttt gttctaattt tccagatgag gaagctgaga ccctaaaagg 360
 ctgaccgggt ccctgatgtg ttacctgtt ctgctactga tccaaactgc agaacttctc 420
 attcatcccc aaggcctcca ggcagtatcc aatggggaat cagctctaaa aggaaccaga 480
 ccaacgtttt ccagccctt cattctgtag cttccctctg tgtgaggaaa ggatagaaat 540
 gttcaggaca tcatcataca ggctcctcat ctacaaagtt ccagtagcag tgacgcctac 600
 acggaagact tggaactgca aacaggctgg ggtcacctca gtgacatctg acgtgtcca 660
 accagaagtt cgatttttgt tctgggggtg aaggaggaaa cagactgtac taaaggacta 720
 aaataatttg tctatactaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaattcccg 780
 cggccgaaag ggaattc 797

<210> 48
 <211> 791
 <212> DNA
 <213> Homo sapien

<400> 48
 caggcgtgag cgtcatgcc tggcagagtt cagcttttat tcacatgttt tccccgaagt 60
 gatttattct tcaaagtaga cagttatgtt ctatagagtg ttttgttttt tctttaagaa 120
 aataatttac ataaacagag attatggtaa acattttaaa tottaggctg ttggttaaatt 180
 ttaatggttt aagcactgtt gggttctctt taattaatat ttgcagaagg agaacatatg 240
 tgtttcactg atatgtatgg tccagaaaaa ttacttaatt ctcaaaaata tgttgcatc 300
 tcatattgtg ttagggaaaa ttccataagt agtctatttt tttttctttt gctgactgtt 360
 aacatccaaa cacctgaatg aaaactgact ctttctgtga ttggtgtgtc acaatattgc 420
 tgtgccgatg ttcacagAAC acttgcggtt ttcgcttcac attgctaaat caaatgtaaa 480
 gccaaatatg tatatttaaat aaatgagaag ttttttttta ttactgaaat ttattctcaa 540
 cgcaaatgta tttttagat gtttcatttg ggagattttg ctttgctta aaacatacga 600
 aataaacctg tcttggtgtc tgccacctc aaaacctctg ttaacttgac atgtagaagg 660
 agttcagaat tctttgataa tgtgtggtt tcacactttt gttgggatta accaaaaata 720
 aaattagagt ccataccact ttgtaaacta atgtgaagtt tcttggtgaa tcataaaagc 780
 tacctgtatg t 791

<210> 49
 <211> 1791
 <212> DNA
 <213> Homo sapien

<400> 49
 gattaatgta gacaaacgtc caggtagcaa ttttggggat aataaatgag ttcacccttt 60
 ttttttcttt ttttccctga gacagagttt gctcttgttg cccaggctgg agtttaaatgg 120
 cacgatcttg gcttaccaca acctctgcct cctgggttca agcaattctc ctgcctcagg 180
 ctcccaagta gctgggatta caggcatgtg ccatcacacc cggctaattt ttgtattttt 240
 agtagagaca gggatatctc atgttggtca ggctgggtc gaactcctga cctcagggtga 300
 tccgcccact tcagcctccc aaagtgtgtg gattacaggc gtgagccgtc atgcctggcc 360
 gagttcagct tttattcaca ttttttcccc gaagtgattt attcttcaaa gtagacagtt 420
 atgtttctata gagtggtttg tttttttttt aagaaaataa tttacataaa cagagattat 480
 ggtaaacatt ttaaattctta ggctgttggg taaatttaat ggtttaagca ctgttggggt 540
 ctctttaatt aatatttgca gaaggagaac atatgtgttt cactgatatg tatgggccag 600
 aaaaattact taattctcaa aaatatgttg cattctcata ttgtgttagg gaaaattcca 660
 taagtagtct attttttttt tcttttctgt actgttaaca tccaaacacc tgaatgaaaa 720
 ctgactcatt tctgtattgg tgtttaaaaa tattgatttg cagatgttca cagaacactt 780
 gcattttttg attcacattg ctaaatacaa tgtaaaggca aatatgtata tttataaat 840
 gagaagtatt tttttattac tgaaatttat tctcaaagca aatgtatttt gtagatgttt 900
 catttgggag attttgcttt gccttaaaac atacaaaata aacctgtctt gtggctctgcc 960
 cacctcaaaa cctctgttaa cttgacatgt agaaggagtt cagaattctt tgataatgtg 1020
 tggttttcac ttttgtttgg attaaacaaa aataaaatta ggtccatag cactttgtaa 1080
 actaatgtga agtttcttgt tgaatcataa agctacctg tatgtacttt ataatttaat 1140
 gttctgttag taaaaattgt cagcatttta tcttttctc ttctcattac attttagtct 1200
 ccaatctttc ccaactctcag cagtcacagt tttgcagagc aaaacatttt tagaaactga 1260
 atatgtgtga gttctatata aaatgaatgt gttagtaaca tccatctgct gatcaaggag 1320
 gcattggatc tgggtactaga aggtgaaatt gattgtagct atcaaagcat tttatcaatg 1380
 taagtcaaga aaaaagaaga aaactgtgaa cctctgatat ttttaacata aaaactgttc 1440
 ccaatgagtg ttctcttgct gattttgtgt taatgttatt gtctatgatt ttttaagctaa 1500
 tgctaataata aaatctaaaa tttcaacatg atgacaacaa ttctgttagc ctgtttttac 1560
 cattaggatg tttttgaaaa cagatgtcat cttagaaatt atatttttaa gtgcaaataa 1620

atcatcctga cttgaaagtc aacacatctt atttttcatt ccgtagtata acagaatatg 1680
 ctgcatttag atacagggtt aatttgccag attttctcaa aatttatatt gctacaactg 1740
 gtttacttaa catgcaattg aattgttatt taaataaatt acatttgatg g 1791

<210> 50
 <211> 526
 <212> DNA
 <213> Homo sapien

<400> 50
 gctacagcat gttgcttctt caacaaattg agttttcagg gaaatccacc agttaattga 60
 attttaattt ttgggtaaaa gtatcaaaca ttcactcttg ccattttctc ctcttaaact 120
 ttattatcta atcaaacata gtttgccata agatgtaata agagatatag ataacggttg 180
 gacattatct gaggatccat ttgtggaact gaacttaata gacacttcag ttcggtgata 240
 cacatcaagt ttctagctgg gttcagttag ctatagtaat gttggctgac tgtctcacag 300
 ctgtcatagt tagcttcttt aagattgtac tgtggctgag gtcaaacaca aatgcttctg 360
 ggtaaccag cggtcttcaa tgtatggttg ttgaccaggc gaacccttta gaagtgattt 420
 ctagtcttaa caagttttgt accttgccct gggcaggcgt ctctgaagcc gaattccatc 480
 aactgcggc gatcgatgg tccgactcgg tgcaacttgg gtaaca 526

<210> 51
 <211> 692
 <212> DNA
 <213> Homo sapien

<400> 51
 acttttcaaa aaggaagaca atttattgga agaatttgta gctggaaaca ctgttcatta 60
 agaaaacatt aagttaccct gagaaagact cttaatatata ccagtgtttt cagggtcccc 120
 tgaccaacgc atggaagagc aacttgtagg ctctagcctt ttaacaatac ctacataaag 180
 aatattttga tctaaaatgt aacttgggtt ctctgccaca ctggtaataa gtcacaacca 240
 agacatctga atgtgatgga gtataagcaa attttgcgtt tatttttaggc tgcctctttt 300
 cttttctaag agaaagagtt tgcagttctt caaagtggc ttggatgaaa cctactgttt 360
 gggcacaaca aaggaatctt ctgtaagtaa actggtagtt ttcttaaaac agtaaacaaa 420
 tttatctggc ctacattctc taaactatta ttatatgcct agaaaataag gcattagtaa 480
 ttcatcattg agcattgcag agcatagaca actgtgtctt tctaatacagg agcataagca 540
 aacattccgg gaaggcgagg gtatttttaa cggtcttat ggtttacagg taacatttga 600

gtccctaata atttcatatt aaaggggttt cccaaggggt tttatacaaa ataatttggg 660
agaacacggg ggagcgccga agagcggggg tc 692

<210> 52
<211> 3979
<212> DNA
<213> Homo sapien

<400> 52
ccctcgagcc gtaccgtcgc ggatttcggc ggcggaaaca tggcggtcgc ggcggggccg 60
gtaacggaga aagtttacgc cgacactggc ctgtattagc gcgtatggcc tcggggccctc 120
gttccccaag gcgtgccgcc tccctgttct cagtcgcagg ctgaagcctt gtctgctctc 180
ctcctttttg gtttggtttt ggaactgaact ccgaggggtg ggagagcgcg ttggtggcga 240
cggccgagtc agatcactat aaacaaaatt tccacaagag aaaatgttga aataggagtt 300
gcggatacat tggatatact ggatgaaata caagcggtta atttttgtaa cgtgagggaa 360
aagccacat tgctggttac atgtgtaaat cactgcgtta ttgctttagt cattgtctct 420
atntagcaat gacaagactg gaagaagtaa atagagaagt gaacatgcat tcttcagtgc 480
ggtatcttgg ctatttagcc agaatcaatt tattggttgc tatatgctta ggtctatacg 540
taagatggga aaaaacagca aattccttaa ttttggtaat ttttattctt ggtctttttg 600
ttcttggaat cgccagcata ctctattact atttttcaat ggaagcagca agtttaagtc 660
tctccaatct ttggtttgga ttcttgcttg gcctcctatg ttttcttgat aattcatcct 720
ttaaaaatga tgtaaaagaa gaatcaacca aatatttgct tctaacatcc atagtgttaa 780
ggatattgtg ctctctggtg gagagaatth ctggttatgt ccgtcatcgg ccactttac 840
taaccacagt tgaatttctg gagcttggtg gatttgccat tgccagcaca actatggttg 900
tggaagaagtc tctgagtgc attttgcttg ttgtagctct ggctatgctg attattgatc 960
tgagaatgaa atctttctta gctattccaa acttagttat ttttgagtt ttgttatttt 1020
tttcttcatt ggaaactccc aaaaatccga ttgcttttgc gtgttttttt atttgctga 1080
taactgatcc tttccttgac atttatttta gtggactttc agtaactgaa agatggaaac 1140
cctttttgtg ccgtggaaga atttgcagaa gactttcagt cgtttttgct ggaatgattg 1200
agcttacatt ttttattctt tccgcattca aacttagaga cactcacctc tggatatttg 1260
taatacctgg cttttccatt tttggaatth tctggatgat ttgtcatatt atttttcttt 1320
taactctttg gggattccat accaaattaa atgactgcc aagagtatat tttactcaca 1380
ggacagatta caatagcctt gatagaatca tggcatccaa agggatgcgc catttttgct 1440

tgatttcaga gcagttggtg ttcttttagtc ttcttgcaac agcgattttg ggagcagttt	1500
cctggcagcc aacaaatgga attttcttga gcatgtttct aatcgttttg ccattggaat	1560
ccatggctca tgggctcttc catgaattgg gtaactgttt aggaggaaca tctgttggtat	1620
atgtatttgt gattcccacc aacttctgca gtcttgatgg tcagccaaca ctgcttcccc	1680
cagaacatgt acaggagtta aatttgaggt ctactggcat gctcaatgct atccaaagat	1740
tttttgcata tcatatgatt gagacctatg gatgtgacta ttccacaagt ggactgtcat	1800
ttgatactct gcattccaaa ctaaaagctt tcctcgaact tcggacagtg gatggacca	1860
gacatgatac gtatattttg tattacagtg ggcacacca tggtagagga gagtgggctc	1920
tagcaggtgg agatacacta cgccttgaca cacttataga atggtggaga gaaaagaatg	1980
gttccttttg ttcccggctt attatcgat tagacagcga aaattcaacc ccttgggtga	2040
aagaagtgg gaaaattaat gaccagtata ttgcagtga aggagcagag ttgataaaaa	2100
cagtagatat tgaagaagct gaccggccac agctaggtga ctttacaaa gactgggtag	2160
aatataactg caactccagt aataacatct gctggactga aaagggacgc acagtgaag	2220
cagtatatgg tgtgtcaaaa cggtaggtg actacactct gcatttgcca acgggaagcg	2280
atgtggccaa gcactggatg ttacactttc ctctgtattac atatcccta gtgcatttgg	2340
caaattgggt atgcggctctg aacctttttt ggatctgcaa aacttgtttt aggtgcttga	2400
aaagattaaa aatgagttgg tttcttcta ctgtgctgga cacaggacaa ggcttcaaac	2460
ttgtcaaatc ttaatttga ccccaaagcg ggatattaat aagcactcat actaccaatt	2520
atcactaact tgccattttt tgtatgctgt atttttattt gtggaaaata ccttgctact	2580
tctgtagctg ctctcacttt gtcttttctt aagtaattat ggtatatata aggcgttggg	2640
aaaaaacatt ttataatgaa agtatgtagg gagtcaaatg cttactgtaa atgcataaga	2700
gacgttaaaa ataacactgc actttcagga atgtttgctt atggtcctga ttagaaagaa	2760
acagttgtct atgctctgca atggtcaatg atgaattact aatgccttat tttctaggca	2820
tataataata gtttagagaa ttagagaccag ataaatttgt ttactgtttt aagaaaacta	2880
ccagtttact tacagaagat tcttttttcc aaacagtagg tttcatccaa gaccatttga	2940
agaactgcaa actctttctc ttagaaaaga aagaggcgag cctaaaataa acgcaaaatt	3000
tgcttatact ccatcacatt cagatgtctt gggtgtgact tattaccagt gtggcagaga	3060
accaagtta cattttagat caaaatattc tttatgtagg tattgttaaa aggctagagc	3120
ctacaagttg ctcttccatg cgttggtcag ggggcctga aaacactggg aatattaaga	3180
gtctttctca gggtaaacta atgttttctt aatgaacagt gttccagct acaaattctt	3240

```

ccaataaatt gtcttccttt ttgaaaagta ctctcataga agaaatttag caatttctcg 3300
ttgactgact cagtctatTT taagtattca gaaaagattt tgatcccat tgagttaatg 3360
ctctgccttg aaaattatTT ttctgatcct tgtttagtgat aacatttttt ttctactgaa 3420
ggtcagagga taggaaacaa gtatttctct tctgggtatac atgtaatgta ttctgtaaaa 3480
aagtattcat attggcaatt ttagttaggc ataatttTgt ggttgtaatt tttaaaactt 3540
agtgtttTgt ctgattaaag caggcactga tcagggtatc tcctaagagg taattcactt 3600
cttattcctt tccaataatt attacattct aaattttcat ctatgagaaa taacaaacaa 3660
gaaggggaata gaattaaatt ggggtataat ctaattctca ttgtttaaat ggtttgcctt 3720
ctcaccattg aagccatttt tttatagcct cagaaagagg aaataatgcc tccaccattt 3780
tctacctggT gacttgaaaa ttgaactttt aagttaggaa gaagttagag tcagggaact 3840
tgtataccac tatctatgca gcattgttat agtctgatta tttctgtgtt ttgaatatga 3900
ttttcctaatt gctctaaata aaattttgtt aaaaattaaa aaaaaaaaaa aaaaaaaaaa 3960
aaaaaaaaaa aatgagcgg 3979

```

```

<210> 53
<211> 478
<212> DNA
<213> Homo sapien

```

```

<400> 53
acctttaact caatttaata taacaagaaa tcgtaaaata cttataacct atcttagaga 60
aatgagtgtt ggttttgaga gttgtttttt aactgaaaga ttatttctag atgggtagtg 120
ctttgtgttg gtttctgtt ccatatattt ccagtcatt ttaattagag aagatactct 180
atggtagaac taaggccttt ctttcttg ccaaagtctt taccctattt aacccttTgt 240
atatttctga ctgtcactg ttcatattat aggggaccag atttgtaata tagaattctc 300
cataacatga atgaaattaa ttctgtccaa gccagcatgg tggcttcata ttaagtagta 360
acagaagtct gaacaattgg ataaatttga cttccaagac agctaaactt ttcaactgca 420
attttaaaaa ctacactaca ctgttatagt taatctgaca aaaatgtcct caaagagt 478

```

```

<210> 54
<211> 1540
<212> DNA
<213> Homo sapien

```

```

<400> 54
gtatcattga tgattactgg aatcgatttt atgtcttttg tattttaatc acttgagtta 60

```

```

atcaaccact ggcaaatccc atttgacaaa gattagcatt gtaaaaaaca gatactgtgg 120
tagatttcta gaaattcatt cacatttaag acttctaaaa tggaataata gccttttggt 180
tttcatgagc atattcgcac ccctatatga attacagcat ttaaagttca aaatcagtaa 240
cttttaatct aggaaattga aaaatattaa gttgcaaagc aaaaaaagggt attttcttga 300
aaatactatt taatgtttta ctagactata ggtagttcct taaggttggt tgacctgaag 360
tggagttggg tttggaagct ggtgccagc tggtatggag tgtgtagttt tgttatgaaa 420
gttctctacc acctacctgt gtgagtgaaca ccaacatcca gatgtcacag ctctccagag 480
ctagtcagaa gagaaatcaa attagtgttt aaacccattt gcatattgac ttgtcagtac 540
ctttaactca atttaataata acaagaaatc gtaaaatact tataacctat cttagagaaa 600
tgagtgtggt ttttgagagt tgttttttta ctgaaagatt atttctagat gggtagtgct 660
ttgtgtggt ttctgcttcc atatatttcc cagtcatttt aattagagaa gatactctat 720
ggtagaacta aggcctttcc tttcttgccc aaagtcttta ccctatttaa ccctttgtat 780
atttctgact gctcactgtt catattatag gggaccagat ttgtaataata gaattctcca 840
taacatgaat gaaattaatt ctgtccaagc cagcatggtg gcttcatatt aagtagtaac 900
agaagtctga acaattggat aaatttgact tccaagacag ctaaaactttt caactgcaat 960
tttaaaaact acactacact gttatagtta atctgacaaa aatgtcttca aagagtactt 1020
tattttattt aaagcatctg ttttaattcaa cctttaataa ttttgcaaag aagggtatgt 1080
gtgtatttta atatagcctg acctgaattt atatgttttt agcttttagta ttttaactttt 1140
tgtaacaaat aaaccttttt tctacaaaca aacacacaca cccccccac ccacacacca 1200
ccacacccca cccacaccaa tccacacccc caccacacaa cacaccagca cccccccagc 1260
ccccgcagct cggcggcaag catacggagc ggggggagtg gtgccgcca atgccaagtg 1320
gcgtcttttg ctaccacag gcggtccgct gagtcgtcga gcgccgaac acggatcgcg 1380
aacgccagcg aacactcagt gcgcgttcca tccacatggg aagcacgagc ggccaccgca 1440
tagccggctt gctctcgat gcgtccctc aatcaacagc tagcacagcg cacgtgatgc 1500
ggcacgctct gccggtcgca cacgcacaac acggggatgc 1540

```

<210> 55

<211> 179

<212> DNA

<213> Homo sapien

<400> 55

gcagggtacat atttaatgta tgtattcaat gatgtaacaa gtaatcaggc aaatatcaac 60

attatagaga ctttaatat gaactggatt ccaacaaaac agttttatta aaataaggca 120
 caatgtttgc atcctcggca aaattcacag caggattggt gtaatataag atcttttgt 179

<210> 56
 <211> 3817
 <212> DNA
 <213> Homo sapien

<400> 56
 ccagcttttag ctatgatgca gcaagcacag cagcccctta ccttcattcc ttcttccttc 60
 ccactttcaa tcaattcatc tattcttttc ctttcttcag actgggcaga gagaaagaaa 120
 aacagcatca gtatcttctc ctaggcccat cgtgcgtagc ttgatggtct tgagccctga 180
 ttgccaggc catgcccacc gggccacaat cggcctcatt tggcatcact ggggatgatg 240
 ggtccccagt gatggcaaag cccccaagta tccctccttt tctcatcacc catctgttgt 300
 ggaagatctg tcacctgggg ttcaactgga tcaggaggga aacagtgggg acccaagaac 360
 agaatggggc tcgtagatat gttctgttgc ccatgcagca cgtaaaaaa tgtccaactt 420
 gccacacct gaaaatcagg cctctgactt cacagaaaat caggtaacgt gggccaggcg 480
 cgggtggctca cgcctgtaat cgcaacactt cgggaggccg aggcgggagg atcataaggt 540
 cacgagttcg agaccagcct ggcaaatagg taaaacctg tctctattaa agatacaaaa 600
 attagccagg tgtggttaga gcctgtagtc ccagctactc gggaggctga ggcaggagaa 660
 tcgcttgaac ctgggagggt gaggttgca ttagccaaga ttgtgctact gcactccagc 720
 ctgggcgaca cagcaagact ccaccttaa aagaaaaaag aaaatcgggt acagcagatc 780
 agaggctgtg ccctttggat gggacacacg cagtcacat ggctctggtc tgatggctca 840
 tactttctgtt tgggatcgct gagattcacc tgtatggagg ccaccacgat ggatgagaag 900
 agggcctcca atcccagggt tcaatacaga cctgaacaga gaactgggag ggggcacccc 960
 tggatccacc tctcctctaa ggccaccct cctgcaceta cctccatccc taacctggg 1020
 ttctactgct ctgccactgc acagatacta cagagcaaaa gggaaccaa tgaagacaga 1080
 tcggaagctc caaaccagtg tggctcacc caaaccagca tgtcttgacg gcatagactt 1140
 tcaccaaacc agatggcacg tgtcaggagc ctgacacaa ctgctgagct cagcccatc 1200
 cccctacaca gaggccaaa ccagcttgca gcttttcag gcactcaatc cacacctgca 1260
 atgtgccagg cgctgcagtc tgtgctggga acaatggtga atgagtaacc tcaaggacag 1320
 tcccaaactc tgccacctcc tctccatctc cattcccact gcggccctgc agccagcca 1380
 cggccccggc cccgctccgg cccccttgc agtcaccagg ctttactct gacccaggg 1440

aacactcagt tctccacaag gtagccagag gggctcttta aaatgtaa	gagggccaggg	1500
gcagtggctc tctcctgtaa tccaacact tcaggaaagc cgggaggaag gatcgcttaa		1560
ggacaggagt tagagaccag cctaagcaac agatccagac cctgtcccta caaaaaataa		1620
ataagctagg tgtgggtggc tacacctttg gtcccagcta ctctagaggc tgaggaagga		1680
ggagattgac tggagcccag gaggttgagg ctgcagttag ccatgactgt gacactgcac		1740
tcctgcctgg gcaacagagt gagaccctgt cttaaaaaaa aacagaaaac atgaccaggc		1800
atgggtggctc acgtctgtca tcccagcact ttgggaagct gaggtgggtg gatcacttga		1860
ggttacagat ttgagaccag cctggccaac atggcgaaac cccgtctcta ctaaaaacac		1920
aaaaattagc tgggcgtggg ggacacacc tgtaatcccg gctactcagg aggctgaggc		1980
aggagaatcg cttgaacca ggaggtggag tttgcagttag gccaagatcg caccactgca		2040
ctccagcctg ggagacagag caagactcta tctcaaaaat aaaaaataaaa aaaaaattgc		2100
gtgcaatttt gtattttcat agtcgtatct ttttaaaggt atcatgattt cagttgtggg		2160
cagggaagtat gtgccttaaa tctctactc tagacccaaa gtttggagag ctatattatt		2220
taataagttg tttgtgacag ccttgttacc tttttcattt gatttgaggg agaaagactg		2280
tgatcctgac agattccttc tcataaaatg gcctaagtgt tatcagtcta ggacttctgg		2340
ggaggggaacc tctaccatgc attctgtccc aggatgtcaa agtcataaga atcagggtcc		2400
cctgaaataa aatcactgaa aagatatgtt ctgttatata ttatttaaaa aatttatctg		2460
gtgccaccaa agaatgacag cagtttctaa ccaacttcat atttatagca tcttatgaag		2520
atattgtaag gcttagcata ttttgccact ggttttcttt gtaatatagg ttgaaagtga		2580
gacatgtttg aatacttttg tatgtaaata tctccattc tttttctatc tcttcttggg		2640
ctatatttac taagaattga tatttaaaaa acagttcact aatgaactct acatattatt		2700
gaacactcac agggcaatat tgatttgggt gctactagac ttttacctaa cattagtctt		2760
tctcaatagt tgttgtaaag gatagtattc aatccagtaa atattaaagt gtattagttt		2820
aatgaagggt atttatatac tgtcatacca caaacctatg gtggaaagaa catctgcatt		2880
caccagaatg tacttgttcc tttggctgtg aataaattgg ataagacttt tttattgtaa		2940
gttccagctg ttggaagata cggggataag attgacattg ctgttgacagt attgcaaaaa		3000
catgactaaa ttggttaatt atgtctaccg cttatgttta agagaatcct ttcactaaact		3060
taaattgtta acattgttgt gatattgaga aagaatatta acctaaacag tcactttaca		3120
acaatcatgt aaagacgtgt gcctgcagtt gaggtttttt gcatttctga gcctgctttg		3180
tattcatgag aaacaaaaac ataatgggag aaaagtttta gataagcagc attgtaagtt		3240

tttgtaaagt ttgggatgtc aaagtattaa caaaggggtac tgaaaacata cttttacttg 3300
 ggtcaaatta ctttttatga tctgatttct taattttctg tatttgaaat cttgcaaatt 3360
 aggaatatct acatctatag ataaataagt aaaacttaat ggtagaaata agtgtaattc 3420
 agcaacatga ttcaacaatt tttatatatta ggataagtta ttgtttatta tattaatatc 3480
 aaatttatat attgccttgt aatgctaaat gctcttaaaa gaatatatgg gctacttcaa 3540
 ttctaccacc ttcttcccc tccccagga cgtacaaaag atcttatatt aaccaatcct 3600
 ctgtgaattt tgccatatca aacattgtgc cttattttaa taaaactgtt ttgttggaa 3660
 ccagttctat attaaagtct ctataatgtt gatatttgcc tgattacttg ttacatcatt 3720
 gaatacatac attaaatatg tactaacatt gactctgttc tagatgcaat ggataaaaga 3780
 taaattggaa aaaaaaagt cgacgcggcc gcaatt 3817

<210> 57
 <211> 265
 <212> DNA
 <213> Homo sapien

<400> 57
 gcaggtactt ctggaataga gagttcaaga aattaggaga aaaatgaact tttgaagctt 60
 tttctttccc ttttttgttt acttcattct cttactcagt tttaaaatgc tggtaatggt 120
 cttttttttc tttttttttt tcttggtgat tttaatgctt tggaaaagat ctcatggttt 180
 tatctccaaa ggaggaaatt aatttgatgc catggaaatt agttttctag tcgtatgcct 240
 tgaatgagtg aagaatttct ttttc 265

<210> 58
 <211> 2184
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (237)..(237)
 <223> a, c, g or t

<400> 58
 cgataatcaa tgttgacctt gcaatttcca tcttgcttat atcccctact ttctattatt 60
 acctttgcct gctacgtgat acttctgtct tgggggttaa agccgttaac ggatgataat 120
 ctctaccga cccagacag cgctcgtgtc acctcatcca catttggggc caccgccgat 180
 acaggtcaat caaaccatcc tgccatgaca acctgggttaa acccggtctc tagccanata 240

caaattactg gtgtggttgc aacacctgta gtcccaacta attcggaagg ctcgccagga	300
aaatcatttg aaccagga gtggggaggt ttcagtgacc gaggagttgc accactgcac	360
tccaacctgg cacagaggtg aagactccgc ctccaaagaa atatatacta ataaagaaca	420
gcagaggaca gtgattttctc ataatacaag ctgaggtgaa gaaatattta aagaaaatga	480
caaatgtata atttcaaatt tagattccag aagcttgcca aacatttggt aaattttctt	540
acaaggaaaa aaaacatcat tggtcagatt caagattttt ttttctttaa tgcacaaaca	600
tataagaaaa aacatctcct ttatcttagg actgaccaac tgtgcctgct ttctttattc	660
tcaacagtct atcacatact cgtactcgtg gcaacaatac tgtgttagat tacgaatgct	720
tgtcttggca aaagagagac aaattcccat cttattactc caaagttcta tgttagtaga	780
ctataacagc aactcaaatt ctgggcattt tagatgtaca gaattagaaa aatgatcaag	840
caaagaagca aatgttctat gaagaaattt ttgaatatca gtttacta aaaggccaaa	900
gtcttaatat taaacatatt tcctttttca cccccaccc ctccccccgc tactgagcat	960
atttatattg acaggtcaca aacaaggggc acgggggctc cactttggga ggccaagggtg	1020
ggcggaccac tttgaggcca ggagtttgac accaacctgg ccaatgtggc gaaaccgtct	1080
ctactaaaaa taaaaaatc agctgggcgt ggtggtgcac acctgcaatc ccagccaccc	1140
ggaggggtgaa gcaggagaat cgcttgaacc caggaggcag aggtttcagt gagccgagat	1200
cgcaccaccg cactccaact gggggacaga gcgagactct gtcccaaaaa gataaaaaata	1260
aataaaaaata aaaataaaaa taaaccaaatt gaatgaagtt tcctccaag tttgtcatct	1320
tcactcttagg aaatagctta aagtttaata aagtttacac atgccaattt tgtgaatatc	1380
aaattcaaca gtttggaac acaagcttct aaataaactg tttcactgtg acagtgtcct	1440
tgagaataca tgccatccag aggtaattct gctttatact cagattcttt ccatacttcc	1500
aaaaaaggat caatattaga cctgtacaac aaattacact cttttacaga aaataataaa	1560
atatccaagt ctctaccaa attttcaaaa aagaggaaaa gtgtaagctt ccagatgaaa	1620
gtttctatag ctttcccaa atttagtacc accatgaaaa agaaattctt cactcattca	1680
aggcatagca ctagaaaact aatttccatg gcatcaaatt aatttcctcc tttggagata	1740
aaaccatgag atctttttcca aagcattaaa atcaccaaga aaaaaaaaaa gaaaaaaaaa	1800
accattacca gcatttttaa actgagtaag agaataaggt aaacaaaaaa gggaaagaaa	1860
aagcttcaaa agttcatttt tctcctaatt tcttgaactc tctattccag aagtacctaa	1920
tgcttttctt aaaagagagg ctttcaattt ttccctatgt cttaaaggctg ctttaagtag	1980
cctaagacca aggacaggag agtgaaaacg aagagggttt tggctctcca aggtgggggt	2040

ggaattgcag ctactgctta gggatatttt ccagtgggtca tctcttcaaa ctccagtgag 2100
 tctcacaac aggggtgcacc agccaatcca agtatccagt atctacaatg caaactgtag 2160
 atactatcca aatcctcgtc aaac 2184

<210> 59
 <211> 449
 <212> DNA
 <213> Homo sapien

<400> 59
 acctcttgcc ttttctgggc ttgcgtttct ctctcttagt ggggtgggat gactttcaat 60
 gactttcaat acttcccctg aaggaagaat gataaggaga aatgtctgtt ctgaggaaag 120
 ggctttgaat tccccagata ctgaacaatt tgtgtttgtg actgatggag aatttcagga 180
 atgaatgaga aagcctttgc gaaactatgc aacagtttac atcagtcctat gtgaacgtat 240
 ttgtctaaaa ctatgagcaa actgaagacc aaattattct cctgttgagg tccgtggatg 300
 gcagatttaa aggaagaac cacaaggct tgcaaagata ggagaggctc catctctaata 360
 gcatgtagaa gctccttacg ggtgtccatc aagagcatag cttggaagcc accatgctgt 420
 gcggaactgc gtcagggcaa atgtacagg 449

<210> 60
 <211> 1441
 <212> DNA
 <213> Homo sapien

<400> 60
 cctggagcag ctggtggagg ccaagtaact ggccaacacc tgcctcttcc aaagtcccca 60
 gcagtggcag gtgtacactg agccctggtt gctggccccg gccggtcaca ttgactgatg 120
 gccaccgcct gacgaatcga gtgcctgtgt gtctacctct ctgaagcctg agcaccatga 180
 ttcccacagc cagctcttgg ctccaagatg agcaccaca ggaagccgac ccaggcctga 240
 ggggccagga acttgctggg tcagatctgt gtggccagcc ctgtccacac catgcctctc 300
 ctgcactgga gagcagtgtt ggcccagccc ctgcggctta ggcttcatct gcttgacat 360
 tgctgtccc agagcccctg tgggtccaca agcccctgtc ctcttccttc atatgagatt 420
 cttgtctgcc ctcatatcac gctgccccac aggaatgctg ctgggaaaag caggacctgc 480
 cagcaggtat gagatctagc ctgctttcag ccatcacctt gccacagtgt ccccggttc 540
 taagcctcca atatcaccct gtgagcctcg cacagctcag ccccaacaca gaggtgagac 600
 caggaataag gccacaagta tctcactttc tctgcagaaa tcaatcttta cttcatcaga 660

44

```

gagaccta aa gcgattctta caaggagctt gctgcaagaa acacgggtcat tcaatcacat 720
tgaggagggt ccacatggca ttgagagggt gctgcccgtt caatgccag cagcagctct 780
ggaaggcagt gctcagcccc atcaccactg tcccgtggat gcctgtgtac ctcttgccctt 840
ttctgggctt gcgtttctct cctctagtgg gtggggatga ctttcaatga ctttcaatac 900
ttcccctgaa ggaagaatga taaggagaaa tgtctgtttt gaggaagggt ctttgaattc 960
cccagatact gaacaatttg tgtttgtgac tgatggagaa tttcaggaat gaatgagaaa 1020
gccttttgca aactatgcaa cagtttacat cagtcagtgt aagtatttgt ctaaaacaga 1080
gcaaaactgaa gaccaaatta ttctcctgtt gaggtccgtg gatggcagat ttaaaggga 1140
gaaccacaaa ggcttgcaaa gataggagag gctccatctc taatgcatgt agaagctcct 1200
tacgggtgcc catcaagagc atagcttggg agccaccatg ctgtgcggaa ctgcgtcagg 1260
gcaaatgtca cagcaggatt tcccacccc agctccatca tcacagacac agagagctgc 1320
aggggaggcc tgcccactgt ttgtgcact ctgccctcct ctggcagcat agatccttag 1380
gtgctcaata aagggtgtgt gtattgaact gaaaaaaaaa aaaaaaaaaa aaaaggcggc 1440
c 1441

```

<210> 61
<211> 514
<212> DNA
<213> Homo sapien

```

<400> 61
acaatgtatg tctgattcac accagggaag tggcacagtg ccctttcttg gatccctac 60
aaagtcaa at tccttagatc ctgagaagtg gagtgcattg gatgccctga aaagggtggg 120
gtgtccctgt gtagcagcca gtaactgatc tgaagggaga ggacttggct ctggtgatgt 180
aacatttcaa gcctctgtgt aattacctag tcttagtctt ttcttctca ttcttagtag 240
agacgtgggg aactttcatg aaaaatgcta attctgactc ctctcagcgt gcaacagatt 300
tggtacactt catccactca gctgcaagat ctgagtgtt ttccagaggtg actggaagag 360
ttctctaata ccctacaaag accatggatc ttgcccactt cagggtgtgt ggctcaaacc 420
tcttaaagtc atcccaggaa aaagtgttga ttgtagtatt ctctcgatgt atgtcaatag 480
aatttatgtc ataataatag taggttctga tgggt 514

```

<210> 62
<211> 2145
<212> DNA
<213> Homo sapien

<400> 62
 ccacctcggtt tgcgtctctt ggggactcta ccgagagacc tctcttttct cccggccatg 60
 gcccgagagt tttttccagg gggctctgaa ccgcagcctc aggttctctg caaggagccc 120
 ctgcttgccc tggggcccg cacccttggt ttccctgaat ccctgggtat aaacctggga 180
 tctctcagag tcccccaag gggaatttct ccccgacccc caacctgga taaggaatca 240
 ctttctgggc ccatttcggg caattccctc aacaatagga atgacccctc ttttcttaaa 300
 accttaccca aacttctgtg cccaccccg gcctcttttt tttttttttt tggataatga 360
 ccttggtttg aggtgcatga gtgaatttta gaaatgaatg tacaatgtat gtctgattca 420
 caccagggga agtggcacag tgccttttct gggatcccta caaagtcaaa ttccttagat 480
 cctgagaagt ggagtgcatt ggatgcctg aaaaggtggg ggtgtccctg tgtagcagcc 540
 agtaactgat ctgaaggag aggacttggc tctgggtgat taacatttca agcctctgtg 600
 taattaccta gtcttagtct tttcttctc attcttagta gagacgtggg gaactttcat 660
 gaaaaatgct aattctgact cctctcagcg tgcaacagat ttgttacact tcatccactc 720
 agctgcaaga tctagagtgc tttcagaggt gactggaaga gttctctaata accctacaaa 780
 gaccatggat ctttgccact tcaggtgctg tggctcaaac ctcttaaagt catcccagga 840
 aaaagtgttg attgtagtat tctctcaatg tatgtaaata gaatttatgt cataataata 900
 gtaggttctg atggtactac ttcttccaa gggagtcact ctactgcacc ctcttgtct 960
 gtgtatacag tgctcaccct tgcaggagca ggaaagtccc tcatctagag ctcaacccca 1020
 gcccttgctc cttaacggtg tgtgtctgtg tagtgagggg ggttggtcaa gcatcccccg 1080
 tcaatgtaga gatgtggcag aaaccggtc acctgttgta ttggtatctg gctccagaaa 1140
 gaaaagtttc attgcttcga cataagaata aattgatgaa tgaagttaa cccagaagag 1200
 gcttcacaaa gaggtcgtgt aagcatctgc ccatgggact cccttccacg caccgtcttt 1260
 ctactaggt gttggggagg acaggagct ggggctggg agggcagtg gaagagggag 1320
 ctttgcttag ggacagggaa aggtgcccc ttctgacag ttgtaggact tttctttccc 1380
 tcctgtcttc cccctcaacc tcctcaaac gtagcctctg gagaacctg actctggcgg 1440
 ctgagggcct acctgtgagt gagctttggg cttccccgcc tgtctttgca caggagcctg 1500
 tgtcaggtg cacctggaca cgctggggg ggagggacat cagcagaggg gggacaggg 1560
 ggcagacacc cccacatccc accaggtagg ctgatgtggc tggaacaaca cccccagatg 1620
 gaatgagtac tcttctcacc ttcccaaata gatccttgag atgtcagcg ctcaccaca 1680
 ctggtcactg tgggtgggta agctgaacac atccttccat gaactgggaa gaggcacaga 1740

46

```

gggagtcaaa atatgccctt ttcttgccctc cattctcctc ccagtccctc ctgtgctgac 1800
at ttgccccca gaggcaggtc ttcttttaaaa tatggaaacg gccagactc catcagcaag 1860
tatttgccctc ccctgggggtt taaagaggtc ttctgggagt cagcaggccc tttttgtggc 1920
ctctttgctg aattgtttct aatccctgac aatgatattt caattcttgg cctctaggga 1980
tggagatgcc atcatcctcc ttaccacct ttcccacgat gaggctaaaa accccgatga 2040
ccagggttcc actctatccc tgacctacat tcgtgttttc tttctttgcc tttaggagtg 2100
gtggctgtgt atcttcagga ctccataaag tagccaccat ctttt 2145

```

<210> 63
<211> 576
<212> DNA
<213> Homo sapien

```

<400> 63
acataccccc agctgcagca gaatatcaat agattctgtt ctcccaggag aagggaagg 60
actgtatcca atcttatctg gggatatgtat ccaaatgacc taagacagtc ttctaataa 120
acacttttgg accgcagggt tcagactctc ctgggggtgga atcttttttt gttacctttc 180
tttctgcctg ctggtttaag tcaggatgca tgcaaggccc acgtctccag tgccccaag 240
tggttgtcta gggtttgcaa gaggacatct agtgatggga gaactcactg cttccagcca 300
ctctgtctat acaccccggt agaaaaatga tctgttgacc agaattttgg cataatttcc 360
tacctttttt tttattaag gggcacagac ttaatcta atctcttccat cataatggtc 420
ttttaatatt ttatgagaga gattcctaaa gtccttcttt agatttaa acctcttatt 480
tttctaacta ttcattaatt aagcattttc atagtccag tgaaatgtaa cgggctgttt 540
ctcgtatctt taaaagtgga gtgcccaggg ctaagt 576

```

<210> 64
<211> 675
<212> DNA
<213> Homo sapien

```

<400> 64
acataccccc agctgcagca gaatatcaat agattctgtt ctcccaggag aagggaagg 60
actgtatcca atcttatctg gggatatgtat ccaaatgacc taagacagtc ttctaataa 120
acacttttgg accgcagggt tcagactctc ctgggggtgga atcttttttt gttacctttc 180
tttctgcctg ctggtttaag tcaggatgca tgcaaggccc acgtctccag tgccccaag 240
tggttgtcta gggtttgcaa gaggacatct agtgatggga gaactcactg cttccagcca 300
ctctgtctat acaccccggt agaaaaatga tctgttgacc agaattttgc cataatttcc 360

```

tacctttttt ttttattaag ggtcacagac ttaatctaata tctcttctct cataatgggc 420
 ttttaactat tttatgagag agattcctaa agtccttctt tagattttaa cacctcttat 480
 ttttctaact attcattaat taagcatttt tcatagtcct agtgaaatgt aacgggcttt 540
 tctcgtatct ttaaaagtgg agtgcccagg gctaagtaca ggagtggctt tggttcacat 600
 ggtgcatatg tagcttgtca tgtgatactt ttttttcag actaaattta ctgtgagcca 660
 ggtgtctctg aatct 675

<210> 65
 <211> 719
 <212> DNA
 <213> Homo sapien

<400> 65
 acacctatta ttctggagat acttgcttct atagatttat tacaatatgt tttataaagt 60
 attttagagt atataatttg tgtttatgtt ccacagaaac atattttata ggagttaatc 120
 ttgactatct aaaggatttg tgaactagtt ccagcttctt ccaataccct tgtccacgag 180
 aagtaaaacta aatcatgtat ctatttcctc tattatcttt attaaataat aagttaatgt 240
 ggctgaata tatacggatt tctgatacta tggcttatta ctgagggaaa aaacaccact 300
 aaactatcct ctaatctgtg taatagatta gctacacttt cttcactagc aagataaaat 360
 aatttcaca ttttctagtt ttactttgta gaaataactc tctgtaattg gactgtattc 420
 aacgaaaact tagtaagttg taattatgcc tcaggtatgt ttctatgcac tgagtgaaga 480
 gtggagataa aaatagaatt tagattttcc tttacttttt aaataggttg ttgcctctta 540
 tatatttatt ctatgatgca aatgtcacta tcttaattcc tcagtttatg tttacagca 600
 cacagtggca cttctatgat tcaaatacat ttgataccct ttgaaatcaa tcagaatact 660
 gcaaaattaa tttttctaaa acatgctttt atcggtattt ctctgttga atcatcagt 719

<210> 66
 <211> 2965
 <212> DNA
 <213> Homo sapien

<400> 66
 ggccgcctt tttttttttt tttttttttt tttttttata cagtatctaa cttatcttta 60
 ttttggaat agctggatta ttacaacctt tgtatcattt gcagggttat tccaatcttt 120
 atagccttgt tgggcttttc tattgaatga tgatcattga cacacgttga aaatattaag 180
 tactcgagaa taatgcctta agcaggagta cttgacacac gtgaaaaatt taacttggtg 240

gcaaacaca	aaagaacaat	ggtaacagta	atgaagccag	aaacctcctt	gcctcccagt	300
aatttgcgac	atatttctac	attttgaagc	cagctagcag	tgtggaacaa	gaaatccgat	360
gcctcaatcc	catttagata	aataaaattt	caagattttc	acaatgatta	ccttcatggc	420
agctgatatt	aaatgagcac	actgaagtat	gctaggcact	gttttaattg	ttttatgtat	480
tatttcatct	ttgcaataaa	tactcattgt	ctacattgta	cagataagga	attgagcgca	540
gaaaagttgt	gacttgctca	agttttcagg	gtagaaaagt	ggcaaagacc	taattctaaa	600
aaggctttat	aattacagat	tttgtgctct	tatcttttgt	tctatactgc	ttggctttca	660
atgttgctc	aaatccctc	ctgatttagc	ccctgctcca	cgcacaaaaa	caatatgcag	720
agttattaac	tagggaagaa	gctgttaatt	tttatgattt	tcctactaca	aagatactca	780
tctatatatt	gaggggtggaa	aattaaaata	gccacagaaa	acagaaatga	gatttcaaaa	840
tataagccag	ttagaatgtc	atagtggcaa	gcaaagttgt	catcaaatag	tcatcaatag	900
tttattatag	caaaatacaa	taaattatat	tttattgaat	tcattaagtg	gcagttaaaa	960
aaggattact	tactgctga	aagtaatgtc	tcgataatgt	ggaaatttta	catatatata	1020
taaaacagtt	ctaatgatca	tacataagaa	gacatttggt	aagacagctt	acataataaa	1080
aacaatttat	acatgggtca	ttgataacca	ccagtatctc	tctttttccc	cggcctttcc	1140
cagttatctg	aagattgctg	cacaaaataa	ttgttttccc	atatatcatt	aatatcaagc	1200
attttgaaga	aattatagta	tctttttttc	tgtatatgaa	aggaattaca	aaatatggag	1260
aagggttgta	tgttgattaa	tggtgaaatg	gggcataata	cttaaccttc	aaaagcctcc	1320
aatgacgcaa	tttttatcac	acagaacata	gggtcaatgg	gaaagagaat	gaagaatgta	1380
gatagaaaat	aatttaggaa	gataacacaa	tagaataggg	tggattgaaa	gggaatacat	1440
gacacttccc	tttgaatgta	tgaatctgag	tgtctatcca	tgtcatgatg	aaaagttctt	1500
gtaagcaatg	ctttggcttt	ttagaaaata	gccctttagt	ttattaagga	aaatttccat	1560
ggatgaggaa	ataatcatat	cattgtcaga	tatttggtat	cactgtcctt	acatcatggg	1620
tctgttagag	aaagattgta	atatgagatt	attttaagtg	ctttcatttg	gaaattgtac	1680
tgatgattca	acaggagaaa	taacgataaa	agcattgttt	tagaaaaatt	aattttgcag	1740
tattctgatt	gatttcaaag	ggtatcaaat	gtatttgaat	catagaagtg	ccactgtgtg	1800
ctgttaaaca	taaactgagg	aattaggata	gtgacatttg	catcatagaa	taaatatata	1860
agaggcaaca	acctatttaa	aaagtaaagg	aaaatctaaa	ttctattttt	atctccactc	1920
ttcactcagt	gcatagaaac	atacctgagg	cataattaca	acttactaag	ttttcgttga	1980
atacagtcca	attacagaga	gttattttcta	caaagtaaaa	ctagaaaatg	tggaaattat	2040

```

tttatcttgc tagtgaagaa agtgtagcta atctattaca cagattagag gatagttag 2100
tgggtgttttt tccctcagta atagaccata gtatcagaaa tccgtatata ttcaggccac 2160
attaacttat tattaataa agataataga ggaaatagat acatgattta gtttacttct 2220
cgtggacaag ggtattggag aaagctggaa ctagtgcaca ataccttag atagtcaaga 2280
ttaactccta taaaatatgt ttctgtggaa cataaacaca aattatatac tctaaaatac 2340
tttataaaac atattgtaat aaatctatag aagcaagtat ctccagaata ataggtgtac 2400
tacttctatg aggtttgttg ttaccactag accaatcctt tgctgggggt ggaaaagaga 2460
aatgttacag cttaggagc tatttttagct attcctggct attcctggct gacagcggag 2520
attcacctgt gaagtcaaaa tacgataagc catagctacc tcagttgtgg ctcagaaagt 2580
ctaacagtat gtccaaaacc accaccccca cccctttcag aacaagtaag ggcccagggt 2640
actgtacctt cagcttgaga accatggctt ggcatataac ttggcacatg tgatatgatc 2700
tcaggaaaaa gactttgctg cacatgggga tataaacaac tacttctaata gccaacctgg 2760
agttaagatc agagcataac tgaaggagac aaagacacaa aaacccttc aaaaaatcag 2820
tgaattcagg agctggtttt tcgaaaagat caacaaaatt gatagaccac cagcaagact 2880
aataaagaag aaaagagaga agaatacaaa agaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2940
aaaaaaaaa aaaaaagatg cggcc 2965

```

```

<210> 67
<211> 303
<212> DNA
<213> Homo sapien

```

```

<400> 67
taaataagata actcagagct taaaaagttt caagttgtta ctttttgggt gctagaagag 60
ttatccttta gtgaccgaa caatttactt atctagaaga atagtgtctg cttagccaac 120
acataacta aagttagaat aataattatt ggcgggtgc agtggtcacg ccggtaaatc 180
ccagcacttt gggaggccaa gcgagcagat catgggtcgg gagtctagaa caggctgggc 240
acatggtgaa cccctcttg taaattcaaa attgcggggg gggggggggg tttccacatc 300
cgg 303

```

```

<210> 68
<211> 405
<212> DNA
<213> Homo sapien

```

```

<400> 68

```


50

acctctgaag cctgaaaaca caggcaataa aattcaccta tttatacttc tttaccaaaag 60
 agaaagcaat ttctgaatac tatctatagt gctaaactaa tgtgaactga ctatcattgc 120
 gataaaagtt tttccttatg atgacaataa agaatgttgc tgaaagactt taatcttgag 180
 agagcagagg taatgtgatg aatgtaattt gctcccagag cctctagaaa ataaagcagt 240
 gtgcaaaata caatatggca ttattattcc agctagggtt tttgcgaaaa taaggttcca 300
 aatgaatgaa gaaaacaaaa tttgatgctc taggttcctt aacttgctat tggacacatg 360
 ggtatttcaa agaaaatcca ccgtgcctac aatacttggt aaagt 405

<210> 69
 <211> 4301
 <212> DNA
 <213> Homo sapien

<400> 69
 gaccgccttt tttttttttt tttttttttt tttatctttt gagactgaac ctttattttc 60
 tgaaaaacag gtatttcata caatctttgc catgttaatg caaatatgca caaagtaggc 120
 atgtatttgt tttccaaaag atgcattatg aacattttca ggaagctggg gtgatttatt 180
 caacttttaa atacaatcac aaaattatat ccatcaggag gcattacaac cttttgtaca 240
 gaaaagccac tatttatata ttgttactaa gacaaggaag attcagttca actcaacttg 300
 ctcttagaat aagggtaaaa agtaaattaa caagtaagtg aagtatgatg ttgttgccac 360
 tgacattaca ggtggaaata taagggaat ttaaaccaga aaaatgacac aataacttta 420
 aagaggagct gaaactttgt caaaaaaga aaaaactatt agcctgtttt caaagaaaaa 480
 cattctaaaa gtgtgcattt cagaacatag aattcttcta agtttaccat cttcaaaaat 540
 cttctaaatt gtatgacact ttacattag cacaacaaac agctttttct aagtctagcc 600
 aagttcccat ggaaggcaaa cgaccctaag tagttcatat ttacagccc ttgaacttat 660
 aaagcttttc tcattaagag tcagttttac ccttctgtaa ataaggatgg tgatactggt 720
 atccaggcct aaaaagcagg aagtgaaca aacccttagg gtttcatgat acagtgaatt 780
 ttccctccc caacgtttgg aaaaaattgg gacacttgct agttcttccc tgtgggaaga 840
 atctttctaa tattaccaa atattgaaaa caaaatctac cttctttaac cttgtatta 900
 gtaattctac ctcttggt tatgggggga aaagtcctag ttttaaattg ctggcatttt 960
 acaagctcaa caagataaaa aattgaacac tggttttcat actctaattt tatgtaaaac 1020
 aaagatgctt aaatgtgca atagtaaagc attcactgat atttgatgta tctgaatagg 1080
 actaacaggc taattgtagg tgctttcata tgaaaataat tgggagaaaa gaagaaccag 1140

ctcctttgat ttcagtactg ccaaaacaag taagcccca gagttaatta caaaaatgta	1200
gaggaaaata ggcccggaag actttgcaat ttaaagtact gcctataata cccagatta	1260
aaaaggaccg aagattgcac atcaagtcta atatttggt gtacttgcac tgcctgtc	1320
ggcattttta aaaaatggct ctttccttaa atttcacacg tcagaacaac cacaattaaa	1380
aaaacaaaaa acaatgcaga taacaccaa cattggacaa tattaagaaa actacttact	1440
aagcttaggt aatagaggca agggttaagg gcagaagtga tggaagtgtt tctttgtatt	1500
acaaaccctc accctaaacc tggggtttgg gctctaaaaa agtaatttag ggaagcagat	1560
caagagcctg attatcagtt ctacatgga aatctcaaca gatttcaaaa gcaactcacc	1620
acaagtgtca gacacaaaag aatcttaata atcatctagc aaactacact aagaactcaa	1680
tgacaaaaaa taatcaatca aacttcaaga cacaccaaag caacagttaa agagaaataa	1740
ctacacaagc taagctctgt agagaaatca actaaaaatt aggcaattct agtaaattaa	1800
agacagcaaa aaacttaaga ttgaacttca atttccaagg gtactttgaa gaggttgttt	1860
gtaggttatt ctgtaccccc tgattcttaa agaaagtga ggagaggggg aggcgacaga	1920
ggaatggagg gaaaggggga ggcaagagag agaaacagga gaggatccct ggttttccac	1980
aggtatttta tcaaaaatgt atgacttcat aactacagca ctctaacttt tctcagtcta	2040
actgctgagg tctcttacga aagtcaaat ggcaagtcac cttgtagaat gacttaacag	2100
aggaagccaa actatacaga ccagggcctt gtccagagtg cagtaaggag caggcttttg	2160
aagtcactta ttacatctgc cttacaacca aacctactgt gtatgccaca ttattcttgg	2220
agggaccact ttaaaacaaa agtgatggat aatttaccag aggcacatgc ttttgaagta	2280
aagctgtttt ttgtttttgc ttgagtaagg tgccaggtag ctctgaagcc tgaaaacaca	2340
ggcaataaaa ttcacctatt tatcttcttt accaaagaga aagcaatttc tgaatactat	2400
ctatagtgtc aaactaatgt gaactgacta tcattgcatg aaaagttttt cttatgatg	2460
acaataaaga atgttgctga aagacttta tcttgagaga gcagaggtaa tgtgatgaat	2520
gtaatttgct ccagagcct ctagaaaata aagcagtgtg caaaatacaa tatggcatta	2580
ttattccagc taggtttttt gcgaaaataa ggttccaaat gaatgaagaa aacaaaattt	2640
gatgcgctag gttccttaac ttgctattgg acacatgggt atttcaaaaa aatccaccgt	2700
gcctacaata cttgttaaag taccaaaaaa gcaactgatac tgaaacacat tctttcatgg	2760
gtacttaacg attacagatt aacgtggcaa tcagaataca aaaaaggccc agagctatgt	2820
ggaatttttt ccttaatacc tttcaagttt gtctgtgaag acggcaacta gaaagtgacc	2880
actgtcctaa tctctaaaat gcagggaaaa tgtactccaa caattttttt ttttatctag	2940

```

tcattttattt ttgacacaga gtctagctct gttgcccagg ctggagtgca gtggcgtgat 3000
cttggtcac tgcaacctct gctcctggg ttcaogccat tctcctgcct cagcctccca 3060
agtagctggg actacaggcg cccgccacca cgcccggcta attttttttg tatttttagt 3120
agagatgggg tttcaccgtg ttagccagga tgggtctcgct ctctaacct catgatccgc 3180
ccgcctcggc ctcccaaagt gctgggatta caggcgtgag ccactgcacc cggccactga 3240
tgacttcttt agactaaaat cctagaagtg tacattattg gctcaaggac ttgaaagctt 3300
ttgtatcaga gttatatgtc tagaaaatgt gcacttcaca ttcccagtggt gtgggcgtga 3360
cttggttca ttagcaaata gcctctctct ttctcctaac ttcattgtgct tccagtgggt 3420
gagatgattt gatgacagct cttaaaggga tgaggtgaca ccctaaaaga aggettgcg 3480
tgatccacag ggagagggga tttgcaactgt tattctctgt ctgcagcccc aagatacaaa 3540
ctatccaagg atgttctgtt ccacaaacag ctctgtccga gtggtaacac cccaaggtcc 3600
cggcccactc accttttctg ccagcagggtt acacttggct gcctctgtgg cctcttttagc 3660
aacacagtca gtgccagaca ctgtaaactc cacataggta gaagggtggga ggggctgaga 3720
agagtacatg aaaatgctct tgaggaaaga tgacttaca tgaaaagggc agagaaggag 3780
agcgggaagt agcctccttc ggggagcacc attcatgcag accatgcgtc ggcccagaag 3840
acatgtccct caatggctta catagggcag ttgtatggg gttgtgttct caagtgttg 3900
tgcttttagag gtaggggttg tgtcttctt gtagggggg caagtatcca tcaccggtcc 3960
ttcagggttg aggtgagttt gcagatagga aaccggaatc agaaggccag cattagactg 4020
ataggagtgg gaaccctgcc tccccccagc cctcactctt gggctgcact tgatacccaa 4080
ggttcagggtt attccaaaat aggggtgggaa agtgaagaga ttaatactc ttccacctgt 4140
ttcatagagc aggtcaaat gaaaagattt agatgaaagt atggaagata tcacttgagc 4200
ttttcttttt taaagtagtt ggagcatatg gacacaataa aatatctgtt atttggtgta 4260
gtctgggggtg gggatgttct gagaatcacc tctgccgaat t 4301

```

```

<210> 70
<211> 299
<212> DNA
<213> Homo sapien

```

```

<400> 70
acctcttccc acctctcatg gatgatggaa gatacaagtg gttattgaaa aatcatatca 60
gtagtttgca aattcagtat aaaccatgaa caggatattt ttctgatagc ggtgagactg 120
caatgtgcta tgtagaaaaa aagctctctc tgccccataa agtgggagtc agaaagaggg 180

```

ctcaagcttc tttatcctct tcagtgccat aaatactgtc acagcaaaaa ggccttcagt 240
 gtctgctggc cagaaacatc tgcccaggca caaatgggcc acaagggcag ggtacctgc 299

 <210> 71
 <211> 1689
 <212> DNA
 <213> Homo sapien

 <400> 71
 taattcttgc ggccagaatt tttttttttt tttttttttt ttgactcgta ttttaatttat 60
 ttagaatctt acaaaaacaa aaaacaaaac aaaaaccacc acaacagaaa aaaaaactaa 120
 atacagaatt tgttacgctt cacggttgat cgtttttact tgcaagagta aataaacctt 180
 gctaaaatgc agccagtaca tttttattgc atgagaccaa atttttcagt tacatatcaa 240
 aaatgattgg gataatcaat tccggacgct tgggtaccgtg ctcccacgaa aggctggatg 300
 cagcaatgca gtatcattgg aacagggcgc acccttcaca cactgatgga cacgctacaa 360
 caggagcgat aacaaaaggg agatttaaaa aagagaacca aatgaaaaca caccaagagc 420
 tgacatccac ctttgtttca agttgtcttt ggatcccatc agatgttgtc tccagatgca 480
 ccatgtcaga ttagtaaagg agaaccatct acacctacat agaaaagtat cttttgctca 540
 gaggaggtag aacctggcca aagttttatt gcagagatac agtgtacctc ttcccacctc 600
 tcatggatga tggaagatac aagtggttat tgaaaaatca tatcagtagt ttgcaaattc 660
 agtataaacc atgaacagga tatttttctg atagcgggtga gactgcaatg tgctatgtag 720
 aaaaaaagct ctctctgccc cataaagtgg gagtcagaaa gagggctcaa gcttctttat 780
 cctcttcagt gccataaata ctgtcacagc aaaaaggcct tcagtgtctg ctggccagaa 840
 acatctgccc aggcacaaat gggccacaag ggcagggtac tggttagggg cccgcagtgg 900
 aaaagccaga caggttctca ccaggggcct gcagagtggc cttcactctg gaggacgcct 960
 gaattacaag tatcaaaaag aaccgcctt tttgggcttc ttcttttcct tgcttagccc 1020
 tgcactaagg ggcagtcttg ctggacggtg ccctgccacg ttgcggcagc ccagatggcc 1080
 gcactgccaa ccacagcacg gcttcccatc gggcgccaga gggagactga gcaaggaggg 1140
 tctgcgtgga ggatgcacac tggaggcaat ctgtgacagg cccaattca cgacaaattt 1200
 agttcccaag actgatccaa atacagaatg cttttacatt tttcaccag totaccaagt 1260
 tgaatagtaa tgaatgaaac ttgtacatga atgaaaaggc cccaagacg ctcacggcaa 1320
 tccttgaaag ttataaagga acattttctt acagggtgcaa aattgtgaac aaatacccaa 1380
 tgtctgcctc ccgggtgctc aacaccatca ttttgatgaa ccacccgggt cttcccaact 1440

cgaattatta aaaacgcctc gatcccgctc tgcttctagg cttacggctg atgtagacaa 1500
 gatccactca ccgatacaag ggtggagagc acagccgttc agggtacccc agggaccagt 1560
 gctgcaatgg gaatctgctt gtctcaccct ccagcagagt cagcctagga ggctccagag 1620
 cagttgcttg gctctctttt ggaggacaat tgttccttaa tgtacattct ctctcttttt 1680
 tttttttog 1689

<210> 72
 <211> 262
 <212> DNA
 <213> Homo sapien

<400> 72
 acgccgctaa atttggggca atttgttaca tagcaatgta tagctcatac aatttctggg 60
 aaaaaaatag tttatttttag aatcattttt gcataatgca agaataataa cattgtcaca 120
 tgaataattt atccttgtat tagtggtgcc aaatatttca ttgtcagtta tatattagct 180
 caaattaaat ttagataat atatattatt attaattgga aagaatgtgt cacattttatc 240
 tttatagctt ttctgtacct gc 262

<210> 73
 <211> 1323
 <212> DNA
 <213> Homo sapien

<400> 73
 agaattatga gtgattcatg tttttctaac ttccctatct gtattaagtg ttctatagtt 60
 tatatttggt actttttaca tcaggaaata gtaaagttat tattttaaac ttatgaacaa 120
 aaaagtaaca agcacatgca agcacagagt tctaccaaatt gcaaaaaatt tcaaataaat 180
 tattcaaatg agacattaac atcacttctg tggtagtttt atatccataa agtctgattc 240
 ttctcctttg aagagatgaa gcttaattct cctcatcctg aaaatgggct ggacttagtg 300
 acttacgtct ttttatttta tttttaattg acaaataata attgtatgta tttatggggg 360
 acaatattat attattatat atgtatacat tatggaatta ttaaatacaag ctaattaaca 420
 tatccataac ctcttataat ttctttgtgg tgagaacatt taaaaatgta ctcttttagc 480
 aattagggac ttacttttaa tacaggaaaa tggaagagac tgtgagactt tgaagtaggt 540
 cataaaagtc actgtggctt cctccttgct ctctcttggg tcaacttgctc tgggggaagt 600
 caactgccat gtcctgagca gccctggaaa gacctacatg atgaagaact gagaccttct 660
 atcaaatgcc agcaggggaat tgaggcctcc tgtcaacagc catttttagaa gtagatcttc 720

55

cagcctcagt caagccttca gatgactgca gccctgtcta atagcttgac cgtaatttca 780
 tgagagacct tcagccagaa aaccaagga aaccattctg gattcctcat cctcagaaac 840
 tgtatgagat aagaagtgtt tgtttagta cgccgctaaa tttggggcaa tttgttacat 900
 agcaatgtat agctcataca atttctggga aaaaaatagt ttatttttaga atcatttttg 960
 cataatgcaa gaatataaac ttgtcacaga ataatttatc cttgttttag gtggtccaaa 1020
 tatttcattg tcagttatat attagctcaa attaaatttt agataatata tattattatt 1080
 aatggtaaag aatgtgtcac atttatcttt atagcttttc tgtacctaat attgtgtctt 1140
 gtgcgtagga tgtgctcaat aaaaattgat tgaataaata agtgaatgaa agaataaatg 1200
 aatgagtga ggaattatct gaaatatttt tataaaattc cccatatgta tgtattactt 1260
 attacaagtc tgggtccata gctgaaaaaa tattaacat tatatatata taaaaaaaaa 1320
 aaa 1323

<210> 74
 <211> 2919
 <212> DNA
 <213> Homo sapien

<400> 74
 agagtttcag ttttggcagc agcgtccagt gccctgccag tagctcctag agaggcaggg 60
 gttaccaact ggccagcagg ctgtgtccct gaagtcagat caacgggaga gaaggaagtg 120
 gctaaaacat tgcacaggag aagtcggcct gagtgggtgcg gcgctcggga cccaccagca 180
 atgctgctct tcgtgctcac ctgcctgctg gcggtcttcc cagccatctc cacgaagagt 240
 cccatatttg gtcccaggga ggtgaatagt gtggaaggta actcagtgtc catcacgtgc 300
 tactaccac ccacctctgt caaccggcac acccggaagt actggtgccg gcagggagct 360
 agagggtggct gcataaccct catctcctcg gagggtacg tctccagcaa atatgcaggc 420
 agggctaacc tcaccaactt cccggagaac ggcacatttg tggatgaacat tgcccagctg 480
 agccaggatg actccgggag ctacaagtgt ggccctgggca tcaatagccg aggcctgtcc 540
 tttgatgtca gcctggaggt cagccagggt cctgggctcc taaatgacac taaagtctac 600
 acagtggacc tgggcagAAC ggtgaccatc aactgccctt tcaagactga gaatgtcaa 660
 aagaggaagt ccttgtacaa gcagataggc ctgtaccctg tgcgtggcat cgactccagt 720
 gggttatgtga atcccaacta tacaggaaga atagccttg atattcaggg tactggccag 780
 ttactgttca gcgttgatc caaccaactc aggcctcagc atgctgggca gtatctctgc 840
 caggctgggg atgattccaa tagtaataag aagaatgctg acctccaagt gctaaagccc 900

gagcccgagc tggtttatga agacctgagg ggctcagtga ccttccactg tgccctgggc 960
 cctgaggtgg caaacgtggc caaatttctg tgccgacaga gcagtgggga aaactgtgac 1020
 gtggtcgtca acaccctggg gaagagggcc ccagcctttg agggcaggat cctgctcaac 1080
 ccccaggaca aggatggctc attcagtgtg gtgatcacag gcctgaggaa ggaggatgca 1140
 gggcgctacc tgtgtggagc ccattcggat ggtcagctgc aggaaggctc gcctatccag 1200
 gcctggcaac tcttcgtcaa tgaggagtcc acgattcccc gcagccccac tgtggtgaag 1260
 ggggtggcag gaagctctgt ggccgtgctc tgcccctaca accgtaagga aagcaaaagc 1320
 atcaagtact ggtgtctctg ggaaggggccc cagaatggcc gctgccccct gctggtggac 1380
 agcgaggggt gggttaaggc ccagtacgag ggccgcctct ccctgctgga ggagccaggc 1440
 aacggcacct tcactgtcat cctcaaccag ctcaccagcc gggacgccgg cttctactgg 1500
 tgtctgacca acggcgatac tctctggagg accaccgtgg agatcaagat tatcgaagga 1560
 gaaccaaacc tcaaggtacc agggaatgtc acggctgtgc tgggagagac tctcaaggtc 1620
 ccctgtcact ttccatgcaa attctcctcg tacgagaaat actggtgcaa gtggaataac 1680
 acgggctgcc aggccctgcc cagccaagac gaaggcccca gcaaggcctt cgtgaactgt 1740
 gacgagaaca gccggcttgt ctccctgacc ctgaacctgg tgaccagggc tgatgagggc 1800
 tgggtactgg gtggagtga gaggggccac ttctatggag agactgcagc cgtctatgtg 1860
 gcagttgaag agaggaaggc agcggggctc cgcgatgtca gcctagcgaa ggcagacgct 1920
 gctcctgatg agaagggtgct agactctggg tttcgggaga ttgagaacaa agccattcag 1980
 gatcccaggc tttttgcaga ggaaaaggcg gtggcagata caagagatca agccgatggg 2040
 agcagagcat ctgtggattc cggcagctct gaggaacaag gtggaagctc cagagcgctg 2100
 gtctccaccc tgggtgcccct gggcctgggt ctggcagtgg gagccgtggc tgtgggggtg 2160
 gccagagccc ggcacaggaa gaacgtcgac cgagtttcaa tcagaagcta caggacagac 2220
 attagcatgt cagacttcca gaactccagg gaatttggag ccaatgacaa catgggagcc 2280
 tcttcgatca ctcaggagac atccctcgga ggaaaagaag agtttgttgc caccactgag 2340
 agcaccacag agaccaaaga acccaagaag gcaaaaaggc catccaagga ggaagccgag 2400
 atggcctaca aagacttcct gctccagtcc agcaccgtgg ccgccgaggc ccaggacggc 2460
 ccccaggaag cctagacggt gtgcgcgct gctccctgca cccatgacaa tcaccttcag 2520
 aatcatgtcg atcctggggg ccctcagctc ctggggaccc cactccctgc tctaacacct 2580
 gcctaggttt ttctactgt cctcagaggc gtgctggctc cctcctcagt gacatcaaag 2640
 cctggcctaa ttgttcctat tggggatgag ggtggcatga ggaggtocca cttgcaactt 2700

57

ctttctgttg agagaacctc aggtacggag aagaatagag gtcctcatgg gtcccttgaa 2760
 ggaagagggg ccaggggtggg agagctgatt gcagaaagga gagacgtgca gcgcccctct 2820
 gcacccttat catgggatgt caacagaatt tttccctcc actccatccc tccctcccgt 2880
 ccttccccctc ttcttctttc cttaccatca aaagatgta 2919

<210> 75
 <211> 27
 <212> PRT
 <213> Homo sapien

<400> 75

Met His Thr Asn Leu Ser Tyr Met Cys Pro Phe Leu Leu Met Ile Phe
 1 5 10 15

Thr Ser Leu Arg Thr Leu Thr Asn Ile Val Cys
 20 25

<210> 76
 <211> 29
 <212> PRT
 <213> Homo sapien

<400> 76

Met Ile Lys Asn Asp Phe Gly Trp Leu Pro Phe Pro Ser Phe Pro Arg
 1 5 10 15

Val Leu Ile Tyr Val Leu His Thr Cys Lys Leu Lys Cys
 20 25

<210> 77
 <211> 38
 <212> PRT
 <213> Homo sapien

<400> 77

Met Ser Leu Ile Lys Lys Ile Ser Thr Thr Gly Leu Phe Cys Leu Gly
 1 5 10 15

Phe Trp Lys His Asn Phe Pro Met His Lys Lys Ala Leu Ser Lys Leu
 20 25 30

Leu Ser Tyr Gly Tyr Asn
 35

58

<210> 78
 <211> 170
 <212> PRT
 <213> Homo sapien

<400> 78

Ala Leu Glu Thr Ala Pro Thr Leu Ala Leu Pro Asp Ser Ser Gln Pro
 1 5 10 15

Phe Ser Leu His Thr Ala Glu Val Gln Gly Cys Ala Val Gly Ile Leu
 20 25 30

Thr Gln Gly Pro Gly Ser Arg Pro Val Ala Phe Leu Ser Lys His Leu
 35 40 45

Asp Leu Thr Val Leu Gly Trp Ser Ser Cys Leu Arg Ala Ala Ala Ser
 50 55 60

Ala Ala Leu Ile Leu Leu Glu Ala Leu Lys Ile Thr Asn Tyr Ala Gln
 65 70 75 80

Leu Thr Leu Tyr Ser Ser His Asn Phe Gln Asn Leu Phe Ser Ser Ser
 85 90 95

His Leu Met His Val Leu Ser Ala Pro Trp Leu Leu Gln Leu Tyr Ser
 100 105 110

Leu Phe Val Glu Ser Pro Thr Ile Thr Ile Ile Pro Gly Arg Asp Phe
 115 120 125

Asn Pro Ala Ser His Ile Ile Pro Asp Thr Thr Pro Asp Pro His Asp
 130 135 140

Cys Ile Ser Leu Ile His Leu Thr Phe Thr Pro Phe Pro His Ile Ser
 145 150 155 160

Phe Phe Pro Val Pro His Pro Asp His Thr
 165 170

<210> 79
 <211> 74
 <212> PRT
 <213> Homo sapien

<400> 79

Met Glu Ser Cys Ser His Arg Cys Leu Asp Leu Ser Leu Ser Leu Ser

59

1 5 10 15

Leu Ser Phe Leu Leu Ser Gln Gln Leu Phe His Arg Gly Ser His Phe
 20 25 30

Glu Arg Leu Lys Tyr Cys Gly Phe Asn Lys Glu Leu Phe Phe Ser Leu
 35 40 45

Ser Lys Ile Leu Ser Glu Arg Asn Lys Met Gly Lys Gly Arg Leu Arg
 50 55 60

Asn Ala Tyr Cys Pro Lys Ser Asn Ser Tyr
 65 70

<210> 80
 <211> 32
 <212> PRT
 <213> Homo sapien

<400> 80

Met His Val Leu Leu Thr Arg Leu Leu Ile Leu Gln Glu Leu Leu Phe
 1 5 10 15

Val Thr Leu Phe Leu Gly Val Met Met Val Leu Val Phe Met Phe Lys
 20 25 30

<210> 81
 <211> 27
 <212> PRT
 <213> Homo sapien

<400> 81

Met Leu Ser Leu Ile Thr Ala Ser Pro Asp Leu Thr His Ser Ala Arg
 1 5 10 15

Ala Glu Gly Lys Pro Arg Met Leu Pro Asp Tyr
 20 25

<210> 82
 <211> 25
 <212> PRT
 <213> Homo sapien

<400> 82

Met Ser His Tyr His Val Ile Ile Cys Ile Asn Ile Ser His Asn Asp
 1 5 10 15

60

Phe His Asn Phe Gln Arg Leu Ile Ser
20 25

<210> 83
<211> 52
<212> PRT
<213> Homo sapien

<400> 83

Met Asp Cys Pro His Ala Ala Pro Thr Ala Cys Cys Gly Met Cys Ser
1 5 10 15

Ser Ser Ser Arg Gly Phe Ser Tyr Ile Leu Thr Leu Leu Asn Thr Val
20 25 30

Met Gly Leu Pro Thr Glu Pro Ser Gln Gly Gly Ala Gln Pro Pro Val
35 40 45

Gly Arg Leu Ala
50

<210> 84
<211> 175
<212> PRT
<213> Homo sapien

<400> 84

Val Leu His Leu Tyr Arg Ser Gly Gln Tyr Leu Gln Asn Ser Thr Ala
1 5 10 15

Ser Ser Ser Thr Glu Tyr Gln Cys Ile Pro Asp Ser Thr Ile Pro Gln
20 25 30

Glu Asp Tyr Arg Cys Trp Pro Ser Tyr His His Gly Ser Cys Leu Leu
35 40 45

Ser Val Phe Asn Leu Ala Glu Ala Val Asp Val Cys Glu Ser His Ala
50 55 60

Gln Cys Arg Ala Phe Val Val Thr Asn Gln Thr Thr Trp Thr Gly Glu
65 70 75 80

Pro Val Gly Glu Ala Leu Pro Arg Glu Met Ala Gly Pro Leu Trp Arg
85 90 95

61

Leu Ile Asp Ser Asp Pro Pro Ser Glu Val Arg Gly Gly Ala Glu Val
 100 105 110

Met Arg Glu Arg Tyr Thr Cys Leu Gln Gly Ser Gln Ile Arg Glu Asn
 115 120 125

Gly Leu Ala Ser Arg Lys Arg Asn Ile Gln Pro Cys Tyr Leu Ser Pro
 130 135 140

Leu Pro Pro Gly Arg Gln Leu Val Phe Phe Lys Thr Gly Trp Ser Gln
 145 150 155 160

Val Val Pro Asp Pro Asn Lys Thr Thr Tyr Val Lys Ala Ser Gly
 165 170 175

<210> 85
 <211> 51
 <212> PRT
 <213> Homo sapien

<400> 85

Met Ser Pro Leu Arg Thr Pro Leu Leu Arg Gly Leu Gln Glu Leu Gly
 1 5 10 15

Glu Glu Trp Lys Ser Ala Lys Arg Ile Thr Ser Phe Ser Lys Ser Met
 20 25 30

Gly Thr Thr Arg Ala Arg Gly Cys Glu Pro Gly Gly Trp Leu Pro Phe
 35 40 45

Thr Gly Leu
 50

<210> 86
 <211> 48
 <212> PRT
 <213> Homo sapien

<400> 86

Met Val Pro Ile Gly Cys Lys Leu Ser Glu Ser Phe His Phe Asp Asn
 1 5 10 15

Leu Ser Tyr His Asp Leu Ile Val Cys Leu Gln Ile Gln Asp Leu Lys
 20 25 30

Ser Phe Leu Ser Gln Ala Trp Lys Glu Leu Leu Tyr Tyr Gln Tyr Cys

62

35

40

45

<210> 87
 <211> 40
 <212> PRT
 <213> Homo sapien

<400> 87

Met Leu Phe Pro Val Ala Val Tyr Ser Tyr Asn Ile Asn Ile Ile Val
 1 5 10 15

Pro Trp Leu Thr Asp Lys Asn Glu Ser Ile Lys Cys Pro Val Ser Glu
 20 25 30

Thr Gln Val Phe Phe Leu His Pro
 35 40

<210> 88
 <211> 34
 <212> PRT
 <213> Homo sapien

<400> 88

Met Ser Trp Ser Leu Pro Ser Leu Lys Asn Leu Ser Cys His Ile Ile
 1 5 10 15

His Val Leu Asn Lys Phe Val Cys Ile Phe Leu Leu Ile Cys Leu Ile
 20 25 30

Ser Ile

<210> 89
 <211> 32
 <212> PRT
 <213> Homo sapien

<400> 89

Met Cys Val Cys Glu Lys Glu Phe Leu Asn Val Phe Tyr Leu Leu Arg
 1 5 10 15

Gly Pro Ser Pro Thr Leu Gly Leu Ser Val Ile Ser Asn His Ile Thr
 20 25 30

<210> 90
 <211> 28
 <212> PRT

63

<213> Homo sapien

<400> 90

Met Lys Pro Gln Cys Cys Lys Phe Thr Val Phe Ala Cys Ser Arg Cys
 1 5 10 15

Phe Val Leu Lys Glu Thr Phe Thr Ile Tyr Leu Leu
 20 25

<210> 91

<211> 111

<212> PRT

<213> Homo sapien

<400> 91

Lys Asp Arg Lys Ser Gly Arg Thr Ala Leu His Leu Ala Ala Glu Glu
 1 5 10 15

Ala Asn Leu Glu Leu Ile Arg Leu Phe Leu Glu Arg Pro Ser Cys Leu
 20 25 30

Ser Phe Val Asn Ala Lys Ala Tyr Asn Gly Asn Thr Ala Leu His Val
 35 40 45

Ala Ala Ser Leu Gln Tyr Arg Leu Thr Gln Leu Asp Ala Val Arg Leu
 50 55 60

Leu Met Arg Lys Gly Ala Asp Pro Ser Thr Arg Asn Leu Glu Asn Glu
 65 70 75 80

Gln Pro Val His Leu Val Pro Asp Gly Pro Val Gly Glu Gln Ile Arg
 85 90 95

Arg Ile Leu Lys Gly Lys Ser Ile Gln Gln Arg Ala Pro Pro Tyr
 100 105 110

<210> 92

<211> 33

<212> PRT

<213> Homo sapien

<400> 92

Met Gly Ile Ser Trp Ser Ala Phe Gly Pro Arg Ile Arg Ile Asp Gly
 1 5 10 15

Ser Pro Pro Pro Cys Leu Leu Pro Thr Pro Pro Leu Leu Pro Leu Cys

64

20

25

30

Leu

<210> 93
 <211> 109
 <212> PRT
 <213> Homo sapien

<400> 93

Arg Asp Glu Ser Pro Glu Pro Gln Arg Pro Ser Trp Ala Arg Ser Arg
 1 5 10 15

His Cys Glu Ala Cys Val Glu Glu Ser Ser Lys Leu Asp Phe Ser Glu
 20 25 30

Phe Gly Ala Lys Arg Lys Phe Thr Gln Ser Phe Met Arg Ser Glu Glu
 35 40 45

Glu Gly Glu Lys Glu Arg Thr Glu Asn Arg Glu Glu Gly Arg Phe Ala
 50 55 60

Ser Gly Arg Arg Ser Gln Tyr Arg Arg Ser Thr Asp Arg Glu Glu Glu
 65 70 75 80

Glu Glu Met Asp Asp Glu Ala Ile Ile Ala Ala Trp Arg Arg Arg Gln
 85 90 95

Glu Glu Thr Arg Thr Lys Leu Gln Lys Arg Arg Glu Asp
 100 105

<210> 94
 <211> 44
 <212> PRT
 <213> Homo sapien

<400> 94

Met Asn Val Asp Thr Phe Leu Glu Asn Ile Tyr Gln Cys Glu Asn Phe
 1 5 10 15

Phe Asn Thr Leu Thr Thr Lys Ile Lys Tyr Ser Leu Ile Ser Leu Phe
 20 25 30

Asn Lys His Gln Asn Asn Val Ser Val Phe Ile Leu
 35 40

65

<210> 95
 <211> 34
 <212> PRT
 <213> Homo sapien

<400> 95

Met Tyr Cys Ile His Phe Tyr Thr Thr Ser Ala Phe Thr Val Thr Asn
 1 5 10 15

Ile Glu Asn Ile Leu Pro Ser Ile Glu Leu His Met Leu Leu Leu Ser
 20 25 30

Val Cys

<210> 96
 <211> 51
 <212> PRT
 <213> Homo sapien

<400> 96

Met His Phe His Gly Ile Val Phe Leu Ser Ser Phe Asn Phe Cys Tyr
 1 5 10 15

Leu Thr Ser Leu Ile Ala Gln Gln Thr Ser Phe Gln Lys Phe Ser Val
 20 25 30

Lys Ala Phe Glu Leu Leu Ile Phe Asp Leu Ile Tyr Ser Gln His Phe
 35 40 45

Ala Thr Phe
 50

<210> 97
 <211> 77
 <212> PRT
 <213> Homo sapien

<400> 97

Asp Ile Tyr Ile Tyr Phe Ala Asp Gly Val Ser Leu Ser Pro Arg Leu
 1 5 10 15

Glu Cys Ser Gly Thr Ile Ser Ala His Cys Asp Leu His Leu Leu Gly
 20 25 30

66

Ser Ser Asp Ser Pro Ala Ser Thr Ser Arg Val Val Gly Thr Thr Gly
 35 40 45

Val Cys His His Ala Trp Thr Val Leu Gly Phe Phe Val Phe Leu Val
 50 55 60

Glu Ile Gly Phe Cys His Leu Asp Gln Ala Asn Leu Glu
 65 70 75

<210> 98
 <211> 36
 <212> PRT
 <213> Homo sapien

<400> 98

Met Ser Val Trp Ser Cys Tyr Gln Pro Val Leu Leu Asn Val Leu Gly
 1 5 10 15

Gln Leu Glu Thr Ile Ile Lys Glu Thr Asp Pro Gly Asp His Gln Ser
 20 25 30

Ser Phe Arg Leu
 35

<210> 99
 <211> 28
 <212> PRT
 <213> Homo sapien

<400> 99

Met Asp Phe Val Lys His Gln Leu Val Asn Ile Phe Lys Phe Ile Ser
 1 5 10 15

Cys Met Ala Leu Val Ser Val Pro Cys Ser Lys Cys
 20 25

<210> 100
 <211> 57
 <212> PRT
 <213> Homo sapien

<400> 100

Met Trp Gly Phe Ile Ala Lys Asn Gly Lys Ile Phe Gly Leu Ile Phe
 1 5 10 15

Cys Lys Phe Ser Leu Cys Leu Gly Asn Ser His Arg Met Trp Arg Asn
 20 25 30

Glu Leu Leu Gly Ser Val Ala Ala Asp Ser Cys Pro Gly Glu Leu Arg
 35 40 45

Ser Gln Asp Arg Gln Arg Lys Thr Val
 50 55

<210> 101
 <211> 61
 <212> PRT
 <213> Homo sapien

<400> 101

Met Phe Lys Ala Ser Ser Ser Pro Thr Tyr Asn Tyr His Leu His Phe
 1 5 10 15

Leu Leu Gln Ser Lys Lys Thr Pro Cys Val Leu Leu Val Ala Leu Ala
 20 25 30

Arg Arg Lys Met Leu Phe Ser Ile Thr Gly Asn Gln Arg Thr Asn Lys
 35 40 45

Asp Asn Pro Ser Leu His Leu Thr Lys Thr Lys Lys Ala
 50 55 60

<210> 102
 <211> 41
 <212> PRT
 <213> Homo sapien

<400> 102

Met Met Thr Pro Ser Leu Phe Lys Phe Leu Tyr Phe Tyr Leu Leu Ser
 1 5 10 15

Arg Asn Tyr Phe Val Gln Cys Glu Asn Val Leu Thr His Gly Arg Arg
 20 25 30

Val Pro Ile Asp Ser Ser Phe Ser Asn
 35 40

<210> 103
 <211> 44
 <212> PRT
 <213> Homo sapien

<400> 103

68

Met Cys Tyr Leu Leu Leu Leu Ile Gln Thr Ala Glu Leu Leu Ile
 1 5 10 15

His Pro Gln Gly Leu Gln Ala Val Ser Asn Gly Glu Ser Ala Leu Lys
 20 25 30

Gly Thr Arg Pro Thr Phe Ser Ser Pro Phe Ile Leu
 35 40

<210> 104
 <211> 48
 <212> PRT
 <213> Homo sapien

<400> 104

Met Arg Ser Ile Phe Leu Leu Leu Lys Phe Ile Leu Asn Ala Asn Val
 1 5 10 15

Phe Cys Arg Cys Phe Ile Trp Glu Ile Leu Leu Cys Leu Lys Thr Tyr
 20 25 30

Glu Ile Asn Leu Ser Cys Gly Leu Pro Thr Ser Lys Pro Leu Leu Thr
 35 40 45

<210> 105
 <211> 109
 <212> PRT
 <213> Homo sapien

<400> 105

Phe Phe Phe Ser Leu Arg Gln Ser Leu Leu Leu Pro Arg Leu Glu
 1 5 10 15

Phe Asn Gly Thr Ile Leu Ala Tyr His Asn Leu Cys Leu Leu Gly Ser
 20 25 30

Ser Asn Ser Pro Ala Ser Gly Ser Gln Val Ala Gly Ile Thr Gly Met
 35 40 45

Cys His His Thr Arg Leu Ile Phe Val Phe Leu Val Glu Thr Gly Tyr
 50 55 60

Leu His Val Gly Gln Ala Gly Leu Glu Leu Leu Thr Ser Gly Asp Pro
 65 70 75 80

Pro Thr Ser Ala Ser Gln Ser Ala Gly Ile Thr Gly Val Ser Arg His

69

85

90

95

Ala Trp Pro Ser Ser Ala Phe Ile His Ile Phe Ser Pro
 100 105

<210> 106
 <211> 46
 <212> PRT
 <213> Homo sapien

<400> 106

Met Val Val Asp Gln Ala Asn Pro Leu Glu Val Ile Ser Ser Phe Asn
 1 5 10 15

Lys Phe Cys Thr Leu Pro Trp Ala Gly Arg Ser Glu Ala Glu Phe His
 20 25 30

His Thr Ala Ala Ile Val Trp Ser Asp Ser Val Gln Leu Gly
 35 40 45

<210> 107
 <211> 24
 <212> PRT
 <213> Homo sapien

<400> 107

Met Arg Trp Ser Gly Gly Pro Glu Asn Thr Gly Asn Ile Lys Ser Leu
 1 5 10 15

Ser Gln Gly Asn Leu Met Phe Ser
 20

<210> 108
 <211> 697
 <212> PRT
 <213> Homo sapien

<400> 108

Met Cys Lys Ser Leu Arg Tyr Cys Phe Ser His Cys Leu Tyr Leu Ala
 1 5 10 15

Met Thr Arg Leu Glu Glu Val Asn Arg Glu Val Asn Met His Ser Ser
 20 25 30

Val Arg Tyr Leu Gly Tyr Leu Ala Arg Ile Asn Leu Leu Val Ala Ile
 35 40 45

70

Cys Leu Gly Leu Tyr Val Arg Trp Glu Lys Thr Ala Asn Ser Leu Ile
 50 55 60

Leu Val Ile Phe Ile Leu Gly Leu Phe Val Leu Gly Ile Ala Ser Ile
 65 70 75 80

Leu Tyr Tyr Tyr Phe Ser Met Glu Ala Ala Ser Leu Ser Leu Ser Asn
 85 90 95

Leu Trp Phe Gly Phe Leu Leu Gly Leu Leu Cys Phe Leu Asp Asn Ser
 100 105 110

Ser Phe Lys Asn Asp Val Lys Glu Glu Ser Thr Lys Tyr Leu Leu Leu
 115 120 125

Thr Ser Ile Val Leu Arg Ile Leu Cys Ser Leu Val Glu Arg Ile Ser
 130 135 140

Gly Tyr Val Arg His Arg Pro Thr Leu Leu Thr Thr Val Glu Phe Leu
 145 150 155 160

Glu Leu Val Gly Phe Ala Ile Ala Ser Thr Thr Met Leu Val Glu Lys
 165 170 175

Ser Leu Ser Val Ile Leu Leu Val Val Ala Leu Ala Met Leu Ile Ile
 180 185 190

Asp Leu Arg Met Lys Ser Phe Leu Ala Ile Pro Asn Leu Val Ile Phe
 195 200 205

Ala Val Leu Leu Phe Phe Ser Ser Leu Glu Thr Pro Lys Asn Pro Ile
 210 215 220

Ala Phe Ala Cys Phe Phe Ile Cys Leu Ile Thr Asp Pro Phe Leu Asp
 225 230 235 240

Ile Tyr Phe Ser Gly Leu Ser Val Thr Glu Arg Trp Lys Pro Phe Leu
 245 250 255

Tyr Arg Gly Arg Ile Cys Arg Arg Leu Ser Val Val Phe Ala Gly Met
 260 265 270

Ile Glu Leu Thr Phe Phe Ile Leu Ser Ala Phe Lys Leu Arg Asp Thr
 275 280 285

His Leu Trp Tyr Phe Val Ile Pro Gly Phe Ser Ile Phe Gly Ile Phe
290 295 300

Trp Met Ile Cys His Ile Ile Phe Leu Leu Thr Leu Trp Gly Phe His
305 310 315 320

Thr Lys Leu Asn Asp Cys His Lys Val Tyr Phe Thr His Arg Thr Asp
325 330 335

Tyr Asn Ser Leu Asp Arg Ile Met Ala Ser Lys Gly Met Arg His Phe
340 345 350

Cys Leu Ile Ser Glu Gln Leu Val Phe Phe Ser Leu Leu Ala Thr Ala
355 360 365

Ile Leu Gly Ala Val Ser Trp Gln Pro Thr Asn Gly Ile Phe Leu Ser
370 375 380

Met Phe Leu Ile Val Leu Pro Leu Glu Ser Met Ala His Gly Leu Phe
385 390 395 400

His Glu Leu Gly Asn Cys Leu Gly Gly Thr Ser Val Gly Tyr Ala Ile
405 410 415

Val Ile Pro Thr Asn Phe Cys Ser Pro Asp Gly Gln Pro Thr Leu Leu
420 425 430

Pro Pro Glu His Val Gln Glu Leu Asn Leu Arg Ser Thr Gly Met Leu
435 440 445

Asn Ala Ile Gln Arg Phe Phe Ala Tyr His Met Ile Glu Thr Tyr Gly
450 455 460

Cys Asp Tyr Ser Thr Ser Gly Leu Ser Phe Asp Thr Leu His Ser Lys
465 470 475 480

Leu Lys Ala Phe Leu Glu Leu Arg Thr Val Asp Gly Pro Arg His Asp
485 490 495

Thr Tyr Ile Leu Tyr Tyr Ser Gly His Thr His Gly Thr Gly Glu Trp
500 505 510

Ala Leu Ala Gly Gly Asp Thr Leu Arg Leu Asp Thr Leu Ile Glu Trp

72

515

520

525

Trp Arg Glu Lys Asn Gly Ser Phe Cys Ser Arg Leu Ile Ile Val Leu
 530 535 540

Asp Ser Glu Asn Ser Thr Pro Trp Val Lys Glu Val Arg Lys Ile Asn
 545 550 555 560

Asp Gln Tyr Ile Ala Val Gln Gly Ala Glu Leu Ile Lys Thr Val Asp
 565 570 575

Ile Glu Glu Ala Asp Pro Pro Gln Leu Gly Asp Phe Thr Lys Asp Trp
 580 585 590

Val Glu Tyr Asn Cys Asn Ser Ser Asn Asn Ile Cys Trp Thr Glu Lys
 595 600 605

Gly Arg Thr Val Lys Ala Val Tyr Gly Val Ser Lys Arg Trp Ser Asp
 610 615 620

Tyr Thr Leu His Leu Pro Thr Gly Ser Asp Val Ala Lys His Trp Met
 625 630 635 640

Leu His Phe Pro Arg Ile Thr Tyr Pro Leu Val His Leu Ala Asn Trp
 645 650 655

Leu Cys Gly Leu Asn Leu Phe Trp Ile Cys Lys Thr Cys Phe Arg Cys
 660 665 670

Leu Lys Arg Leu Lys Met Ser Trp Phe Leu Pro Thr Val Leu Asp Thr
 675 680 685

Gly Gln Gly Phe Lys Leu Val Lys Ser
 690 695

<210> 109

<211> 36

<212> PRT

<213> Homo sapien

<400> 109

Met Thr Gly Lys Tyr Met Glu Ala Glu Thr Ser Thr Lys His Tyr Pro
 1 5 10 15

Ser Arg Asn Asn Leu Ser Val Lys Lys Gln Leu Ser Lys Pro Ala Leu

73

20

25

30

Ile Ser Leu Arg
35

<210> 110
<211> 21
<212> PRT
<213> Homo sapien

<400> 110

Met Tyr Val Phe Asn Asp Val Thr Ser Asn Gln Ala Asn Ile Asn Ile
1 5 10 15

Ile Glu Thr Leu Ile
20

<210> 111
<211> 130
<212> PRT
<213> Homo sapien

<400> 111

Met Pro Thr Gly Pro Gln Ser Ala Ser Phe Gly Ile Thr Gly Asp Asp
1 5 10 15

Gly Ser Pro Val Met Ala Lys Pro Pro Ser Ile Pro Pro Phe Leu Ile
20 25 30

Thr His Leu Leu Trp Lys Ile Cys His Leu Gly Phe Asn Trp Ile Arg
35 40 45

Arg Glu Thr Val Gly Thr Gln Glu Gln Asn Gly Ala Arg Arg Tyr Val
50 55 60

Leu Leu Pro Met Gln His Val Lys Lys Cys Pro Thr Cys Pro His Leu
65 70 75 80

Lys Ile Arg Pro Leu Thr Ser Gln Lys Ile Arg Tyr Ser Gly Pro Gly
85 90 95

Ala Val Ala His Ala Cys Asn Arg Asn Thr Ser Gly Gly Arg Gly Gly
100 105 110

Arg Ile Ile Arg Ser Arg Val Arg Asp Gln Pro Gly Lys Asp Gln Pro
115 120 125

Gly Lys
130

<210> 112
<211> 31
<212> PRT
<213> Homo sapien

<400> 112

Met Leu Val Met Val Phe Phe Phe Phe Phe Phe Phe Leu Val Ile Leu
1 5 10 15

Met Leu Trp Lys Arg Ser His Gly Phe Ile Ser Lys Gly Gly Asn
20 25 30

<210> 113
<211> 107
<212> PRT
<213> Homo sapien

<400> 113

Pro Leu Pro Pro Leu Leu Ser Ile Phe Ile Leu Thr Gly His Lys Gln
1 5 10 15

Gly Ala Arg Gly Leu His Phe Gly Arg Pro Arg Trp Ala Asp His Leu
20 25 30

Arg Pro Gly Val Ala His Gln Pro Gly Gln Cys Gly Glu Thr Val Ser
35 40 45

Thr Lys Asn Thr Lys Ile Ser Trp Ala Trp Trp Cys Thr Pro Ala Ile
50 55 60

Pro Ala Thr Arg Arg Val Lys Gln Glu Asn Arg Leu Asn Pro Gly Gly
65 70 75 80

Arg Gly Phe Ser Glu Pro Arg Ser His His Arg Thr Pro Thr Trp Gly
85 90 95

Thr Glu Arg Asp Ser Val Pro Lys Arg Ala Lys
100 105

<210> 114
<211> 58
<212> PRT

75

<213> Homo sapien

<400> 114

Met Leu Leu Met Asp Thr Arg Lys Glu Leu Leu His Ala Leu Glu Met
1 5 10 15

Glu Pro Leu Leu Ser Leu Gln Ala Phe Val Val Leu Pro Phe Lys Ser
20 25 30

Ala Ile His Gly Pro Gln Gln Glu Asn Asn Leu Val Phe Ser Leu Leu
35 40 45

Ile Val Leu Asp Lys Tyr Val His Met Asp
50 55

<210> 115

<211> 46

<212> PRT

<213> Homo sapien

<400> 115

Met Ser Asp Ser His Gln Gly Ser Gly Thr Val Pro Phe Leu Gly Ser
1 5 10 15

Pro Thr Lys Ser Asn Ser Leu Asp Pro Glu Lys Trp Ser Ala Trp Asp
20 25 30

Ala Leu Lys Arg Trp Gly Cys Pro Cys Val Ala Ala Ser Asn
35 40 45

<210> 116

<211> 45

<212> PRT

<213> Homo sapien

<400> 116

Met His Pro Asp Leu Asn Glu Gln Ala Glu Arg Lys Val Thr Lys Lys
1 5 10 15

Asp Ser Thr Pro Gly Glu Ser Glu Pro Cys Gly Pro Lys Val Phe Ile
20 25 30

Arg Lys Thr Val Leu Gly His Leu Asp Thr Tyr Pro Arg
35 40 45

<210> 117

76

<211> 45
 <212> PRT
 <213> Homo sapien

<400> 117

Met Trp Lys Leu Phe Tyr Leu Ala Ser Glu Glu Ser Val Ala Asn Leu
 1 5 10 15

Leu His Arg Leu Glu Asp Ser Leu Val Val Phe Phe Pro Ser Val Ile
 20 25 30

Asp His Ser Ile Arg Asn Pro Tyr Ile Phe Arg Pro His
 35 40 45

<210> 118
 <211> 60
 <212> PRT
 <213> Homo sapien

<400> 118

Gln Pro Gly Val Lys Ile Arg Ala Ala Leu Lys Glu Thr Lys Thr Gln
 1 5 10 15

Lys Pro Leu Gln Lys Ile Ser Glu Phe Arg Ser Trp Phe Phe Glu Lys
 20 25 30

Ile Asn Lys Ile Asp Arg Pro Pro Ala Arg Leu Ile Lys Lys Lys Arg
 35 40 45

Glu Lys Asn Gln Lys Glu Lys Lys Lys Lys Lys Lys
 50 55 60

<210> 119
 <211> 32
 <212> PRT
 <213> Homo sapien

<400> 119

Met Cys Pro Ala Cys Ser Arg Leu Pro Thr His Asp Leu Leu Ala Trp
 1 5 10 15

Pro Pro Lys Val Leu Gly Phe Thr Gly Val Thr Thr Ala Pro Gly Gln
 20 25 30

<210> 120
 <211> 41
 <212> PRT

77

<213> Homo sapien

<400> 120

Met Pro Tyr Cys Ile Leu His Thr Ala Leu Phe Ser Arg Gly Ser Gly
1 5 10 15

Ser Lys Leu His Ser Ser His Tyr Leu Cys Ser Leu Lys Ile Lys Val
20 25 30

Phe Gln Gln His Ser Leu Leu Ser Ser
35 40

<210> 121

<211> 105

<212> PRT

<213> Homo sapien

<400> 121

Met Gln Gly Lys Cys Thr Pro Thr Ile Phe Phe Phe Ile Ala Ser Phe
1 5 10 15

Ile Phe Asp Thr Glu Ser Ser Ser Val Ala Gln Ala Gly Val Gln Trp
20 25 30

Arg Asp Leu Gly Ser Leu Gln Pro Leu Pro Pro Gly Phe Thr Pro Phe
35 40 45

Ser Cys Leu Ser Leu Pro Ser Ser Trp Asp Tyr Arg Arg Pro Pro Pro
50 55 60

Arg Pro Ala Asn Phe Phe Cys Ile Phe Ser Arg Asp Gly Val Ser Pro
65 70 75 80

Cys Ala Pro Gly Trp Ser Arg Ser Pro Asn Leu Met Ile Arg Pro Pro
85 90 95

Arg Pro Pro Lys Val Leu Gly Leu Gln
100 105

<210> 122

<211> 38

<212> PRT

<213> Homo sapien

<400> 122

Met Gly Gln Arg Glu Leu Phe Phe Tyr Ile Ala His Cys Ser Leu Thr

78

1 5 10 15

Ala Ile Arg Lys Ile Ser Cys Ser Trp Phe Ile Leu Asn Leu Gln Thr
20 25 30

Thr Asp Met Ile Phe Gln
35

<210> 123
<211> 15
<212> PRT
<213> Homo sapien

<400> 123

Met Gln Glu Tyr Lys His Cys His Met Asn Asn Leu Ser Leu Tyr
1 5 10 15

<210> 124
<211> 764
<212> PRT
<213> Homo sapien

<400> 124

Met Leu Leu Phe Val Leu Thr Cys Leu Leu Ala Val Phe Pro Ala Ile
1 5 10 15

Ser Thr Lys Ser Pro Ile Phe Gly Pro Glu Glu Val Asn Ser Val Glu
20 25 30

Gly Asn Ser Val Ser Ile Thr Cys Tyr Tyr Pro Pro Thr Ser Val Asn
35 40 45

Arg His Thr Arg Lys Tyr Trp Cys Arg Gln Gly Ala Arg Gly Gly Cys
50 55 60

Ile Thr Leu Ile Ser Ser Glu Gly Tyr Val Ser Ser Lys Tyr Ala Gly
65 70 75 80

Arg Ala Asn Leu Thr Asn Phe Pro Glu Asn Gly Thr Phe Val Val Asn
85 90 95

Ile Ala Gln Leu Ser Gln Asp Asp Ser Gly Arg Tyr Lys Cys Gly Leu
100 105 110

Gly Ile Asn Ser Arg Gly Leu Ser Phe Asp Val Ser Leu Glu Val Ser
115 120 125

Gln Gly Pro Gly Leu Leu Asn Asp Thr Lys Val Tyr Thr Val Asp Leu
 130 135 140

Gly Arg Thr Val Thr Ile Asn Cys Pro Phe Lys Thr Glu Asn Ala Gln
 145 150 155 160

Lys Arg Lys Ser Leu Tyr Lys Gln Ile Gly Leu Tyr Pro Val Leu Val
 165 170 175

Ile Asp Ser Ser Gly Tyr Val Asn Pro Asn Tyr Thr Gly Arg Ile Arg
 180 185 190

Leu Asp Ile Gln Gly Thr Gly Gln Leu Leu Phe Ser Val Val Ile Asn
 195 200 205

Gln Leu Arg Leu Ser Asp Ala Gly Gln Tyr Leu Cys Gln Ala Gly Asp
 210 215 220

Asp Ser Asn Ser Asn Lys Lys Asn Ala Asp Leu Gln Val Leu Lys Pro
 225 230 235 240

Glu Pro Glu Leu Val Tyr Glu Asp Leu Arg Gly Ser Val Thr Phe His
 245 250 255

Cys Ala Leu Gly Pro Glu Val Ala Asn Val Ala Lys Phe Leu Cys Arg
 260 265 270

Gln Ser Ser Gly Glu Asn Cys Asp Val Val Val Asn Thr Leu Gly Lys
 275 280 285

Arg Ala Pro Ala Phe Glu Gly Arg Ile Leu Leu Asn Pro Gln Asp Lys
 290 295 300

Asp Gly Ser Phe Ser Val Val Ile Thr Gly Leu Arg Lys Glu Asp Ala
 305 310 315 320

Gly Arg Tyr Leu Cys Gly Ala His Ser Asp Gly Gln Leu Gln Glu Gly
 325 330 335

Ser Pro Ile Gln Ala Trp Gln Leu Phe Val Asn Glu Glu Ser Thr Ile
 340 345 350

Pro Arg Ser Pro Thr Val Val Lys Gly Val Ala Gly Ser Ser Val Ala

80

355

360

365

Val Leu Cys Pro Tyr Asn Arg Lys Glu Ser Lys Ser Ile Lys Tyr Trp
 370 375 380

Cys Leu Trp Glu Gly Ala Gln Asn Gly Arg Cys Pro Leu Leu Val Asp
 385 390 395 400

Ser Glu Gly Trp Val Lys Ala Gln Tyr Glu Gly Arg Leu Ser Leu Leu
 405 410 415

Glu Glu Pro Gly Asn Gly Thr Phe Thr Val Ile Leu Asn Gln Leu Thr
 420 425 430

Ser Arg Asp Ala Gly Phe Tyr Trp Cys Leu Thr Asn Gly Asp Thr Leu
 435 440 445

Trp Arg Thr Thr Val Glu Ile Lys Ile Ile Glu Gly Glu Pro Asn Leu
 450 455 460

Lys Val Pro Gly Asn Val Thr Ala Val Leu Gly Glu Thr Leu Lys Val
 465 470 475 480

Pro Cys His Phe Pro Cys Lys Phe Ser Ser Tyr Glu Lys Tyr Trp Cys
 485 490 495

Lys Trp Asn Asn Thr Gly Cys Gln Ala Leu Pro Ser Gln Asp Glu Gly
 500 505 510

Pro Ser Lys Ala Phe Val Asn Cys Asp Glu Asn Ser Arg Leu Val Ser
 515 520 525

Leu Thr Leu Asn Leu Val Thr Arg Ala Asp Glu Gly Trp Tyr Trp Cys
 530 535 540

Gly Val Lys Gln Gly His Phe Tyr Gly Glu Thr Ala Ala Val Tyr Val
 545 550 555 560

Ala Val Glu Glu Arg Lys Ala Ala Gly Ser Arg Asp Val Ser Leu Ala
 565 570 575

Lys Ala Asp Ala Ala Pro Asp Glu Lys Val Leu Asp Ser Gly Phe Arg
 580 585 590

81

Glu Ile Glu Asn Lys Ala Ile Gln Asp Pro Arg Leu Phe Ala Glu Glu
595 600 605

Lys Ala Val Ala Asp Thr Arg Asp Gln Ala Asp Gly Ser Arg Ala Ser
610 615 620

Val Asp Ser Gly Ser Ser Glu Glu Gln Gly Gly Ser Ser Arg Ala Leu
625 630 635 640

Val Ser Thr Leu Val Pro Leu Gly Leu Val Leu Ala Val Gly Ala Val
645 650 655

Ala Val Gly Val Ala Arg Ala Arg His Arg Lys Asn Val Asp Arg Val
660 665 670

Ser Ile Arg Ser Tyr Arg Thr Asp Ile Ser Met Ser Asp Phe Glu Asn
675 680 685

Ser Arg Glu Phe Gly Ala Asn Asp Asn Met Gly Ala Ser Ser Ile Thr
690 695 700

Gln Glu Thr Ser Leu Gly Gly Lys Glu Glu Phe Val Ala Thr Thr Glu
705 710 715 720

Ser Thr Thr Glu Thr Lys Glu Pro Lys Lys Ala Lys Arg Ser Ser Lys
725 730 735

Glu Glu Ala Glu Met Ala Tyr Lys Asp Phe Leu Leu Gln Ser Ser Thr
740 745 750

Val Ala Ala Glu Ala Gln Asp Gly Pro Gln Glu Ala
755 760